Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Text Mining Based Approach for Customer Sentiment and Product Competitiveness Using Composite Online Review Data

Zhanming Wen [1], Yanjun Chen [1], Hongwei Liu [1] and Zhouyang Liang [2,*]

[1] School of Management, Guangdong University of Technology, Guangzhou 510520, China
[2] Center of Campus Network & Modern Educational Technology, Guangdong University of Technology, Guangzhou 510006, China
* Correspondence: liangzy@gdut.edu.cn

Abstract: We aimed to provide a realistic portrayal of customer sentiment and product competitiveness, as well as to inspire businesses to optimise their products and enhance their services. This paper uses 119,190 pairs of real composite review data as a corpus to examine customer sentiment analysis and product competitiveness. The research is conducted by combining TF-IDF text mining with a time-phase division through the k-means clustering method. The study identified 'quality', 'taste', 'appearance packaging', 'logistics', 'prices', 'service', 'evaluations', and 'customer loyalty' as the commodity dimensions that customers are most concerned about. These dimensions should therefore serve as the primary entry point for improving the commodity and understanding customers. A review of customer feedback reveals that the composite reviews can be divided into three time stages. Furthermore, the sentiment expressed by customers can become increasingly negative over time. The product competitiveness based on the composite review can be characterised by four regional quadrants, such as 'Advantage Area', 'Struggle Area', 'Opportunity Area', and 'Waiting Area', and merchants can target these areas to improve product competitiveness according to the dimensional distribution. In the future, it will also be possible to take customer demographics into account in order to gain a deeper understanding of the customer base.

Keywords: composite reviews; text mining; sentiment analysis; product competitiveness

check for updates

Citation: Wen, Z.; Chen, Y.; Liu, H.; Liang, Z. Text Mining Based Approach for Customer Sentiment and Product Competitiveness Using Composite Online Review Data. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 1776–1792. https://doi.org/10.3390/jtaer19030087

Academic Editor: Ercan Oztemel

Received: 19 February 2024
Revised: 20 June 2024
Accepted: 3 July 2024
Published: 15 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Online shopping is the primary method of shopping for customers. Customer reviews can provide insight into the quality of the purchased goods and the customer's overall experience. They represent a significant source of information for potential customers, exerting a considerable influence on their purchasing decisions [1]. In the absence of clear and reliable information, potential customers are inclined to rely on prior customer experience feedback and product reviews to reduce the decision risk associated with online shopping information uncertainty [2]. The customer base is heterogeneous, comprising individuals with varying personal characteristics, levels of knowledge, and economic status [3,4]. Furthermore, the shopping interests and preferences of customers are dynamic and may undergo change over time. The academic community has been engaged in the study of customer behaviour, with a particular focus on the dynamic nature of customer interests. This is due to the differences between customer demand preferences and actual purchases, as well as the phenomenon of emotional drift in the distribution of emotions over time. It is crucial to note that these variables can be effectively managed with the appropriate approach. In the context of a highly competitive business market, it is of paramount importance to accurately identify customers' changing interests and emotional trends. This enables businesses to enhance their products and services, strengthen their core competitiveness, and increase customer satisfaction and purchase conversions. The objective of this study is to provide a realistic portrayal of customer sentiment and product

competitiveness. This is achieved by motivating merchants to optimise their products, improve their service and increase purchase conversion rates. To this end, the paper examines the textual differences between initial and additional reviews and performs sentiment analysis based on the composite review data of fresh fruit on e-commerce platforms. The analysis identifies the dimensions of customer concerns and discovers the competitive advantages of the products. This enables the acquisition of clear and profound management insights.

2. Related Research

2.1. Composite Online Reviews and Sentiment Analysis

The composite review function, which has been implemented by numerous e-commerce platforms, has significantly enhanced the review system. It enables customers to submit two reviews of the same shopping experience, the initial review and an additional review. The term 'initial reviews' is employed to describe product reviews provided by customers after they have received the product. In contrast, 'additional reviews' are product reviews provided by customers who have used the product for a certain period of time after completing the initial reviews. The additional review provides a highly accurate and comprehensive representation of the product's features and attributes. It includes valuable intelligence information [5], which has attracted the attention of numerous scholars [6,7]. Zhao et al. investigated the impact of management response to online reviews on customer satisfaction, based on additional reviews [8]. Chen et al. examined the interaction effect of online composite reviews and time pressure on customer information persuasiveness [9]. Shi et al. found that additional reviews differed from initial reviews in terms of content length, number, and affective intensity. Furthermore, the additional reviews indicated that the customer's perception of the product evolved over time, exhibiting a more nuanced attitude in the later stages of the customer experience [10]. In a study by Wang et al., two types of reviews were compared in terms of their perceived usefulness by customers [11]. Zhang and colleagues examined the influence of alterations in the emotional valence of composite reviews on customers' willingness to adopt information [12]. Zhou et al. demonstrated that the combination of reviews with different emotional tendencies can differentially affect customer purchase decisions by investigating the impact of additional reviews on customer information adoption [13]. The findings of researchers indicate that the direct sale of goods can be influenced by additional reviews, including the number of reviews and the time interval between them [14]. It is evident that sentiment analysis in customer reviews has become a subject of considerable interest to scholars [15–17]. This highlights the importance of additional reviews in understanding customer purchase behaviour. The current approach to sentiment analysis is characterised by the existence of multiple paths, including those based on sentiment dictionaries, rules and machine learning. Among the aforementioned methods, the sentiment dictionary-based approach is notable for its simplicity and intuitiveness, as well as its suitability for use in the absence of complex models and large amounts of data. Furthermore, it exhibits excellent expandability, which has contributed to its widespread adoption by scholars engaged in the field of sentiment analysis. This study employs a sentiment dictionary-based method for the purpose of carrying out sentiment analysis.

In conclusion, the majority of research in this field compares the value of initial and additional reviews, emphasising their impact on customers' decision-making behaviours, including temporal, emotional, and textual characteristics. Furthermore, both merchants and research scholars have also focused on the length and number of reviews. It is regrettable that previous studies have not examined the temporal changes in customer preferences and product competitiveness. Nevertheless, it is crucial to recognise that additional reviews can provide invaluable insights and that there is a time-series dependence effect on customers' purchasing decisions. In light of the time-varying initial and additional reviews, it can be posited that they can assist merchants in accurately depicting changes in customers' attitudes and interest drift. This is significant in identifying customer prefer-

ences, improving product competitiveness, and increasing the conversion rate of customer purchases.

2.2. Customer Interest Mapping

Online reviews can be anonymous, massive, and disparate, placing an information burden on customers who require access to content of interest quickly. It is typical for customers to rely on a small subset of reviews, typically between five and ten, when making shopping decisions [18]. Consequently, content-based recommender systems have emerged as a valuable tool for enhancing customer loyalty and conversion rates. This is achieved by developing a customer interest model to represent customer preferences and actively recommending information of interest to customers [19]. Current research in the field of e-commerce is focused on the representation of the interests of customers [20,21]. Various methods exist for representing customer interest, including the keyword list method [22], ontological representation [23], topic representation [24], and vector space model representation [25]. The vector space model-based representation is the most commonly used due to its strong computability and operability [25]. Previous research has predominantly concentrated on the construction of customer preference models [26], designing recommendation algorithms [27], customer interest prediction [28] and detecting interest drift [29]. However, this approach fails to distinguish between customers' interests in different time stages and lacks a comprehensive representation of the distribution of customers' interests and their variations over time. The paper presents a clear and concise division of the customer interest stage into three distinct categories: same-day, short-term, and long-term. The vector space model is employed to provide an accurate description of the customer's preferences at different time stages. In addition, both qualitative and quantitative measurements of the specific interest value are provided. This demonstrates that the customer is highly interested in the information and that there is a strong correlation between the customer and the interest in the information. Consequently, merchants are able to discern the customer's personal interest during different time stages, which allows them to provide a more tailored service and achieve mutual benefit.

2.3. Review-Based Product Competitiveness Analysis

The concept of product competitiveness is defined as the combined capabilities and relative competitive advantages demonstrated by two or more products in competition [30,31]. Those merchants who are able to identify the characteristic dimensions of customer concerns from feedback information and to formulate effective marketing strategies will gain a competitive advantage in the market. Previous research has sought to ascertain how commodities can gain a competitive advantage in the market by analysing customer feedback. Xu et al. developed a graphical model to visualise and compare the relationship between products and customer reviews using data from the Amazon e-commerce platform. This approach assists enterprises in identifying potential crises and implementing risk management strategies with precision and confidence [32]. Wu et al. present an innovative framework based on hybrid text mining and visualisation tools for competitiveness analysis and recommendations for product improvement [33]. In their research, Kim et al. employed text mining and sentiment analysis on social media data from Twitter and Facebook to compare the relationship between social opinion and merchandise sales performance [34]. The objective of the research was to predict the performance of the market, evaluate the advantages of competing products, and implement timely adjustments to market strategies. Xiao et al. conducted a sentiment analysis of online reviews on product characteristics, which enabled companies to confidently identify primary and secondary competitors [35]. Xu et al. identified the key factors that lead to customer satisfaction or dissatisfaction based on online reviews of hotels, providing companies with the necessary insights to support their business marketing activities through online reviews [36]. Zhao et al. proposed a product competitiveness model based on mutual information and quantile regression [37]. The

model integrates a multitude of online sources in order to analyse product competitiveness with a high degree of confidence.

Previous studies have predominantly focused on sentiment analysis and text mining of online reviews from the perspective of customer satisfaction. However, there has been a paucity of studies that have attempted to analyse product competitiveness based on customer interest preference drift. The contribution of this paper is to employ a real-world corpus of composite reviews, combining text mining techniques such as TF-IDF with k-means clustering to analyse customer sentiment and product competitiveness. The approach employs time-stage segmentation to identify shifts in customer interests and preferences over time while also monitoring changes in customer sentiment and offering insights for enhancing product competitiveness. Furthermore, it enhances the depth and breadth of the existing research, thereby making a valuable contribution to the field.

3. Research Design

3.1. Data Collection and Pre-Processing

This paper presents a comprehensive analysis of online composite reviews of fresh fruit category products on the Jingdong e-commerce platform, which is a well-known and prominent entity in the Chinese e-commerce landscape. The research dataset comprises online reviews of fresh fruit products delivered by Jingdong Logistics from January 2019 to January 2022, collected using Python. The categories encompass five representative categories of fresh fruits. The research analysed 17,319 pairs of composite reviews, including customer names, product IDs, initial review times and content, additional review times and content, and other information. In terms of data pre-processing, this study employs the Python drop_duplicates method for filtering and deduplication, with the objective of eliminating duplicates and data with missing fields. The process yielded 11,919 pairs of valid comment data. Subsequently, once the valid comment data has been obtained, the key detail is to successively use the HIT deactivation word list processing, employing Jieba disambiguation to disambiguate and deactivate the valid comment data. This is performed in order to obtain the specific corpus and to prepare the data for the subsequent textual and sentiment analysis.

3.2. Time-Stage Processing

The extant literature unequivocally demonstrates the existence of a temporal pattern in customer reviewing behaviour [13]. By comprehending the customer's behavioural patterns during different time stages, merchants can effectively identify customer demand preferences and review motivations. This allows for the reasonable distribution of enterprise resources and the provision of satisfactory customer service. This research characterises the distribution of the time interval sequence between the initial and additional reviews, as well as the number of additional reviews. The time interval is calculated for composite reviews, and the analysis explores customer demand preferences and sentiment trends over different time stages. The optimal clustering k value is determined using dichotomous k-means and the elbow rule [37–39]. Equation (1) demonstrates how the intra-cluster error variance *SSE* accurately reflects the cohesion of the clusters.

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} dist(x, c_i)^2 = \|x - c_i\|_2^2, i = 1, 2, \cdots, \mathrm{k} \quad (1)$$

In this context, k is the number of clusters, C_i represents the ith cluster, c_i represents the C_i centre of the cluster, and x represents a sample from that cluster.

The contour coefficients are calculated for each point in the dataset, and the resulting average value is employed as the overall contour coefficient for the current cluster. A

higher value of this coefficient indicates superior model performance, with the range being between −1 and 1. Equation (2) displays the contour coefficient for i.

$$L(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{2}$$

In this equation, $a(i)$ represents the average distance between vector i and other points in the same cluster, while $b(i)$ represents the minimum of the average distance between vector i and all non-similar points belonging to different clusters.

3.3. Calculation of Text Feature Weights

A multitude of algorithms exist for text feature weighting, with the most prevalent being the word frequency-inverse document frequency (TF-IDF), information gain, chi-square test, and other methods. Among the aforementioned techniques, TF-IDF is capable of capturing the importance of individual words within a corpus of documents. The process of computing and comprehending TF-IDF is relatively straightforward. Furthermore, it serves to reduce the impact of common words on document similarity, thereby enhancing the visibility of the key information within a given document. Consequently, TF-IDF is regarded as a general and efficient text feature weighting algorithm to assess the importance of a word in a particular document across the corpus. The algorithm ensures that the greater the frequency of a word in a document and the lower the frequency of that word in other documents in the corpus, the greater the contribution of that word. In this paper, we utilise TF-IDF to calculate the degree of importance of the service feature dimensions. The main calculation formula is shown in Equations (3)–(5):

$$TF_{(i,j)} = \frac{N(t_j, i)}{S_i} \tag{3}$$

$$IDF_{(j)} = log(\frac{m}{N(t_j) + 1}) \tag{4}$$

$$TF - IDF = P_{ij} = \frac{N(t_j, i)}{S_i} * \log\left(\frac{N}{N(t_j) + 1}\right), i = 1, 2, \cdots, \mathrm{m}; j = 1, 2, \cdots, \mathrm{n} \tag{5}$$

In this equation, the symbol N represents the frequency of occurrence; $N(t_j, i)$ represents the number of times the t_j word appears in document i; S_i represents the total number of words in document i; m represents the total number of documents; $N(t_j)$ represents the number of documents in the corpus that contain that word.

The matrix W, which represents the sum of the importance of each word in the corpus, is given by Equation (6). The data is normalised through a linear function, as demonstrated by Equation (7):

$$W = \left(\sum_{i=1}^{m} P_{i1}, \sum_{i=1}^{m} P_{i2}, \ldots, \sum_{i=1}^{m} P_{im}\right)^T \tag{6}$$

$$W_i^* = \frac{W_i - \min W}{\max W - \min W}, i = 1, 2, \cdots, \mathrm{n} \tag{7}$$

In this equation, $\sum_{i=1}^{m} P_{i1}$ represents the sum of the importance of the words in the corpus, of which W_i is set as the value of the ith row and the first column of the matrix W in order to differentiate the weights of the various dimensions.

In order to gain the customer's attention ($I^{(r)}$) to dimension r in the overall evaluation, it is crucial to focus on indicators with larger weights. Equation (8) clearly demonstrates how the weights of word sets with the same feature dimensions are accumulated. A greater frequency of keywords on the same page indicates a higher level of discussion. However, subjectivity can easily affect the accuracy of the summarised word set when manually

screening words with varying feature dimensions. The manual screening of words with varying feature dimensions may be subject to bias, which could potentially affect the accuracy of the summarised word set. To address this issue, the actual word frequency of feature dimensions can be recalculated by combining the number of words, the word frequency sum, and the weight sum of the feature dimension word set. In order to calculate the actual word frequency ($H^{(r)}$), it is necessary to take the average number of occurrences of each word of the same dimension, as shown in Equation (9).

$$I^{(r)} = \sum_{j \in v_r} W_i^* = \sum_{j \in v_r} \frac{W_i - \min W}{\max W - \min W}, r = 1, 2, \cdots, 8 \tag{8}$$

$$H^{(r)} = \frac{I^{(r)} * \sum\limits_{j \in v_r} \sum\limits_{i=1}^{m} N(t_j, i)}{L^{(r)}} \tag{9}$$

In this equation, W_i^* represents the weight of the word t_i in the corpus; $j \in v_r$ represents the set of words t_j belonging to dimension r; $\sum\limits_{i=1}^{m} N(t_j, i)$ represents the word t_j frequency in the corpus; and $L^{(r)}$ represents the number of words associated with dimension r.

3.4. Calculation of Emotional Intensity

This paper confidently presents a classification of the Chinese emotion vocabulary of Dalian University of Technology into seven emotion categories based on emotion attributes: music, good, sadness, surprise, fear, evil and anger, comprising a total of 27,079 emotion words [40,41].

The interpretation of customer reviews can be challenging when based solely on emotion vocabulary, as negative words and degree adverbs are often used to express emotional tendencies. Negation vocabulary and degree adverbs have a significant impact on the emotional tone and intensity of a review. To accurately understand the customer's emotional tendency, we define the intensity value by constructing a negation dictionary and a degree adverb dictionary.

In order to calculate sentiment scores, it is necessary to adhere to the following rules: if only sentiment words are present, the sentiment intensity value should be used directly. In the event that a negative sentiment is present, the sentiment tendency should be reversed for an odd number of negative sentiments and maintained unchanged for an even number of negative sentiments. Multiply the sentiment intensity value by the adverb of degree if adverbs of degree are present, i.e., $E = d_i Q_g,\ i = 1, 2, 3, 4$.

Combining a negative word with an adverb of degree in the format 'adverb of degree + negative word' reverses the affective tendency, considering the effect of the adverb of degree, i.e., $E = (-1)^n d_i Q_g,\ n = 1, 2, \cdots, n,\ i = 1, 2, 3, 4$. Similarly, if the format is 'negative word + adverb of degree', the affective tendency is reversed by 0.5, taking into account the effect of the adverb of degree, i.e., $E = (-1)^n d_i Q_g * 0.5,\ n = 1, 2, \cdots, n;\ i = 1, 2, 3, 4$. The sentiment score for each review is presented in Table 1, calculated by weighting the number of lexicons included in the corpus according to their number. Table 1 also includes a list of the lexicons and their corresponding sentiment intensity values.

Table 1. Emotional intensity values for selected words.

Dictionary and Number	Meaning	Selected Words	Emotional Intensity Value
Emotional lexicon (27,079)	Literal definition	Fresh, just right, fair-minded, truthful, tasty, tasty, like……	$Q_g, g = 0, 1, \cdots 9$
Negative word (125)	Opposite sentiment	No, bad, didn't, never, nearly, no longer, never, well……	$f = -1$
Adverb of degree (260)	Strong	Most, extremely, extremely, excessively, exceptionally, too……	$d_1 = 2$
	Slightly stronger	More, more and more, doubly, extra, very, quite, better……	$d_2 = 1.5$
	Comparisons	Comparatively, yet, not too, not much, simply, really, as much as possible……	$d_3 = 1.2$
	Marginally	Slightly, a little, somewhat, slightly, only……	$d_4 = 0.8$

In conclusion, the data obtained from the comments was validated by removing duplicates from the composite data set. Subsequently, the k-means algorithm was then employed to perform the clustering analysis and complete the division of the comment time stage. The TF-IDF algorithm and sentiment word classification were employed to conduct a competitive analysis based on the feature statistics of this paper and sentiment analysis based on sentiment scores, respectively. The ultimate objective of this research was to inspire merchants to optimise their products and improve the conversion rate of their services and purchases, based on the aforementioned analyses (Figure 1).

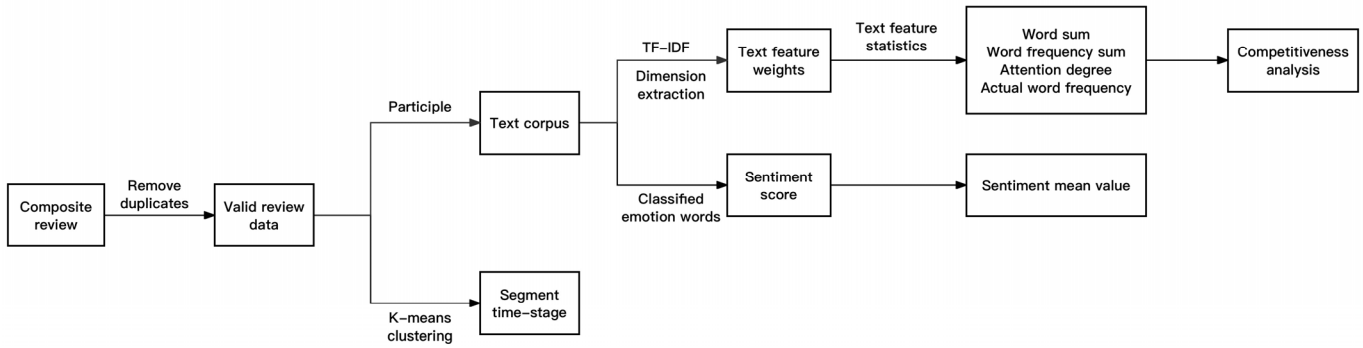


Figure 1. Diagram of analysis methodology.

4. Empirical Analysis and Discussion

4.1. Text Analysis of Composite Review

4.1.1. High-Frequency Word Extraction and Dimensioning

In this paper, we present a list of high-frequency words and their frequencies in online reviews (Table 2). The frequency of occurrence of high-frequency words are significantly higher in the preliminary reviews than in the additional reviews. The initial reviews exhibit the highest word frequency of 2625, with 14 words having a frequency greater than 1000. In contrast, the highest word frequency in the additional reviews is 2040, with only six words having a frequency greater than 1000. The additional reviews emphasise the fresh food characteristics, using high-frequency words such as 'delicious', 'fresh', 'texture', and 'flavour'. The initial and additional reviews both focus on customer attention, although there are significant differences. The two texts exhibit a 76% similarity in high-frequency words, but the distribution and sorting of these words are inconsistent. The initial evaluation highlights words such as 'evaluation', 'Jingdong', 'customer', 'content', and 'unfilled', while the additional reviews focus on terms such as 'delicious', 'apple', 'Jingdong', 'very', and 'good'. The online composite reviews are influenced by the varying demands and preferences of the customers, which are reflected in the different review types. This highlights the diversity of the customer base in terms of review dimensions. Our research has identified eight feature dimensions that customers consider when purchasing fresh products: quality, taste, appearance packaging, logistics, prices, service, evaluations, and customer loyalty (Table 3). To improve these commodities and better understand our customers, we confidently recommend focusing on these dimensions. In fact, the diversity of high-frequency words in the additional reviews may be indicative of differences in the degree to which the product fulfils the customers' needs.

Table 2. The frequencies of the top 50 high-frequency words.

Initial Reviews						Additional Reviews					
Order	**High-Frequency Word**	**Word Frequency**	**Order**	**High-Frequency Word**	**Word Frequency**	**Order**	**High-Frequency Word**	**Word Frequency**	**Order**	**High-Frequency Word**	**Word Frequency**
1	evaluation	2625	26	this time	627	1	tasty	2040	26	shopping	403
2	Jingdong	2437	27	stuff	596	2	apple	1647	27	once	401

Table 2. *Cont.*

Initial Reviews						Additional Reviews					
Order	High-Frequency Word	Word Frequency	Order	High-Frequency Word	Word Frequency	Order	High-Frequency Word	Word Frequency	Order	High-Frequency Word	Word Frequency
3	customer	2425	28	moisture	555	3	Jingdong	1592	28	customer service	381
4	content	2357	29	dissatisfied	547	4	very	1465	29	very sweet	375
5	not filled	2297	30	size	540	5	nice	1378	30	favorable review	374
6	nice	2152	31	grape	522	6	purchase	1025	31	Easy to use	369
7	very	2150	32	not	522	7	fresh	971	32	deserve	367
8	apple	2067	33	character	511	8	fruits	888	33	we	366
9	tasty	1807	34	shopping	496	9	goods	822	34	speed	361
10	fruits	1623	35	merchants	488	10	stuff	816	35	logistics	358
11	fresh	1593	36	once	477	11	no	812	36	character	353
12	package	1490	37	somewhat	475	12	really	765	37	receive	350
13	no	1260	38	speed	474	13	taste	730	38	boxful	338
14	purchase	1115	39	autumn moon	471	14	flavour	646	39	next time	335
15	taste	946	40	fresh food	444	15	quality	631	40	merchants	328
16	received	895	41	deserve	443	16	package	628	41	recommend	326
17	banana	854	42	feel	436	17	like	561	42	express delivery	324
18	flavour	854	43	very sweet	433	18	banana	547	43	autumn moon	320
19	quality	824	44	function	432	19	price	522	44	few days	310
20	express delivery	819	45	deliver	432	20	special	494	45	after	306
21	like	788	46	comparisons	421	21	not	489	46	evaluation	298
22	special	745	47	pears	418	22	this time	456	47	negative feedback	293
23	really	697	48	dispatch	410	23	moisture	451	48	deliver goods	292
24	price	684	49	self-operated	394	24	really	443	49	supermarket	290
25	logistics	682	50	already	383	25	grape	418	50	dissatisfied	289

Table 3. Initial and additional reviews on the high-frequency word sets.

Feature Dimension	Initial Reviews	Additional Reviews
Quality	fresh, quality, stuff, moisture, character, autumn moon, fresh food, product, Korla, light meal, red Fuji, fruit, Xinjiang, spoilt fruit, Yantai, health, pulps, Luochuan, Aksu, premium, Hebei, variety, goods, ample, authentic, very full, rose, juice, fruit, full, sunshine	quality, fresh, moisture, premium, spoilt fruit, not spoilt fruit, light meal, ample, authentic, junk, red Fuji, Aksu, Yantai, autumn months, fresh, fruit, pulp, goods, baby, stuff
Taste	yummy, texture, flavour, very sweet, tastes, sweetness, icing sugar, crunchy-sweet delicate, not sweet, tasty, unpalatable	yummy, texture, flavour, very sweet, crunchy-sweet, icing sugar, sweetness, tasty, sweet, delicate
Appearance Packaging	packaging, a box, inside, net weight, many, single fruit, size, open, very large, found, looking, 2 kg, 10 kg, two boxes, two, even, completely, 100 g, intact, each, 120 g, three boxes, picture, obviously, several, see, intact, large fruit, half, bump, all, seemingly	a box, inside, obviously, net weight, packaging, seemingly, single fruit, size, large fruit, very large, found, looking, 2 kg, two boxes, 100 g, intact, 120 g, 10 kg, three boxes, picture, several, see, intact, half, completely, bump
Logistics	received, expressage, logistics, speed, delivery, shipping, self-support, soon, expressage, delivery, days, delivered, today, time, super-fast, hours, home delivery	packaging, box, very large, two boxes, one pound, half, much, several, each, three boxes, picture, intact, appearance, inside, obviously, looking, see, many, great, size, open, basically, completely, terrible, intact
Prices	price, event, supermarket, affordable, cheap, value for money, deals	price, affordable, cheap, value for money, deals, events, supermarkets
Service	evaluation, customer service, problem, service, after-sales, merchant, attitude, timely, several, service attitude, review, brand, platform, mall	customer service, after-sales, merchant, problem, attitude, evaluation, review, service, 48 h, processing, shop to buy, service attitude, order, brand
Evaluations	unfilled, good, like, satisfied, worth, feel, good, bad, feel, disappointment, everyone, assured, friends, convenient, children, trust, pleasant, almost, suitable, description, experience, protection, cost-effective	good review, negative review, satisfied, experience, bad, disappointment, five stars, unforgettable, thrilled, pleasant, regret, almost, best, not enough, good, family, trust, reassurance, genuine, like, child, really, quite
Customer loyalty	buy, shopping, once, compare, first, many times, before, recommend, repurchase, later, will, finally, order, second time, don't, continue, choose, hope, should, guarantee, before, last time	once, worth, next, recommend, later, first, repurchase, continue, recommend, come back, second, guarantee, often, definitely, as always, order, hope, everyone, many times, before, finally, compare, at first, purchase

4.1.2. Segmentation of Review Slots

In this paper, we utilized dichotomous k-means clustering to partition the review data into time stages. The range of k values was set to [2,10], and we executed the k-means

algorithm to obtain the *SSE* value of the intra-cluster variance for each k value. This allowed us to characterise the distribution of the 'initial review–additional review' time interval sequence and the number of additional reviews. The clustering results (Figure 2) indicate a significant improvement in the degree of aberration when the inflexion point k value is set to 3, with a profile coefficient of 0. The principle of clustering performance evaluation of the profile coefficient divides temporal clustering into three phases: the same day (within 0 days), short-term (within 1–4 days), and long-term (within 5–180 days). The number of corresponding reviews is 6955, 2441, and 2524, respectively. The outcome of having the greatest number of reviews posted on the day of receipt is consistent with customer behaviour. The majority of customers who post a review typically evaluate it immediately after receipt. The comparable number of reviews in the short-term and long-term phases, on the other hand, may be related to the specific type of product purchased by the customer and the duration of its use.

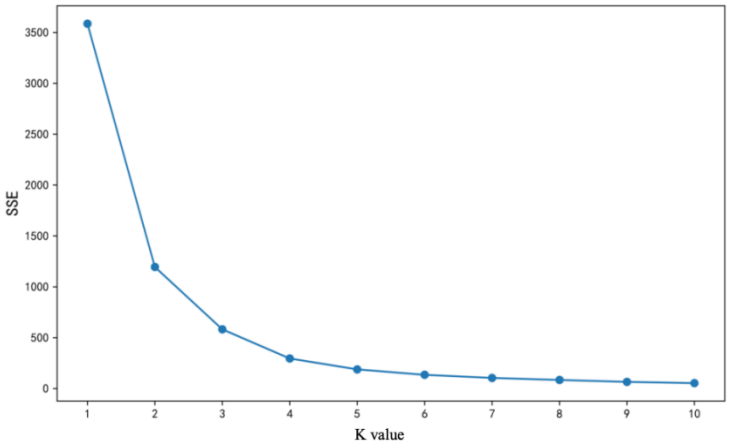


Figure 2. Time-stage clustering results.

4.1.3. Textual Characterisation of Reviews

The results of the cluster analysis presented above provide the foundation for this research. In order to identify customer preferences and differences in composite reviews, the review set is divided into three time stages. The high-frequency word sets of the feature dimensions of each time period are extracted separately, and the screened high-frequency words are derived from the vocabulary in Table 3. The algorithm for weighting text features based on TF-IDF ($F1$ = 0.930, Precision = 0.933) calculates the sum of words, sum of word frequencies, attention level, and actual word frequency for each feature dimension across various review types (Table 4). Furthermore, in this study, the accumulated weight of the word set of the same feature dimension is defined as the attention degree. This is employed to reflect the degree of customer attention to various dimensions of the product. At the same time, this study also defines the actual word frequency of dimension keywords appearing on the same review page as the discussion degree. This is employed to reflect the amount of customer discussion on various dimensions. Attention and discussion intensity are important indicators for depicting the competitiveness of dimensions discussed later in the text. The feature dimensions significantly improve the ranked discrimination of word frequencies, as evidenced by the comparison between the sum of word frequencies and the actual word frequencies. The data indicates that logistics and distribution were the most frequently discussed topics, with a greater variety of related words. Nevertheless, the actual word frequency was not higher than that of taste and flavour.

Table 4. Text feature statistics of dimensions.

Feature Dimension	Review Type	Word Sum	Word Frequency Sum	Attention Degree	Actual Word Frequency
Quality	Initial reviews	79	9342	16.23	1812
	Additional reviews	57	6246	12.29	1343
Taste	Initial reviews *	35	5848	9.51	1519
	Additional reviews *	34	5166	9.17	1395
Appearance Packaging	Initial reviews *	91	8219	14.41	1223
	Additional reviews	66	4433	10.23	666
Logistics	Initial reviews *	45	5908	9.32	1163
	Additional reviews	42	3368	6.78	558
Prices	Initial reviews	20	2287	3.72	408
	Additional reviews	18	1536	2.92	278
Service	Initial reviews	28	3025	4.63	523
	Additional reviews	36	2368	5.07	326
Evaluations	Initial reviews	60	10,197	14.24	2308
	Additional reviews *	80	7349	14.68	1348
Customer loyalty	Initial reviews	63	8883	12.96	1766
	Additional reviews *	83	6720	14.43	1177

Note: Initial reviews * and additional reviews * indicate that the feature dimensions of actual word frequencies do not match the ordering of the sum of word frequencies.

The statistics of the most dominant feature dimensions in different periods are shown in Table 5. The comparison reveals that the dominant feature dimensions in the two types of reviews appear to change with the review period. In the initial evaluation, the number of advantageous feature dimensions will shift from logistics and customer loyalty to longer-term feature dimensions such as taste, prices and evaluations. As the number of advantageous feature dimensions increases, the focus of the reviews shifts from logistics, service and evaluations to long-term quality and prices. This shift is accompanied by a gradual decrease in the number of advantageous feature dimensions. The shift in the dimensions of advantageous characteristics can be attributed to the fact that customers do not have a profound experience of the day's service. Instead, their evaluation content is mostly intuitive feelings, such as logistics. Over time, however, they can pay greater attention to the intrinsic nature of fresh fruits and problematic feedback, such as the quality, taste, and so on.

Table 5. The most advantageous interest preference statistics in three time stages.

Type of Review	The Same Day	Short-Term	Long-Term
Initial reviews	Logistics, Customer loyalty (two feature dimensions)	Quality, Appearance Packaging, Service (three feature dimensions)	Taste, Prices, Evaluations (three feature dimensions)
Additional reviews	Logistics, Service, Evaluations (three feature dimensions)	Taste, Appearance Packaging, Customer loyalty (three feature dimensions)	Quality, Prices (two feature dimensions)

This research visualises the distribution of feature dimensions in composite reviews across different time stages in Figure 3. For each period, the percentage of actual word frequency of feature dimensions in composite reviews is calculated. A higher percentage of actual word frequency of feature dimensions corresponds to a higher area range. The results show that the proportion of reviews related to the 'quality' dimension exhibited fluctuations across different time stages, with percentages of 18.06%, 23.39%, and 17.90%. This indicates that customers prioritize quality in the short term. Furthermore, customer

preferences are subject to change over time, with a particular focus on different aspects during different time stages. The preferences of customers for different types of reviews exhibit distinct differences, indicating a clear pattern of succession and complementarity. For instance, the dimension of 'quality' increases by 22.81% in the initial review, but decreases by 20.85% in the subsequent review. The initial review focuses on the 'service' and 'appearance packaging', while the additional review covers 'taste', 'flavour', 'appearance packaging', and 'customer loyalty'. These dimensions collectively demonstrate the customer preferences across various types of reviews.

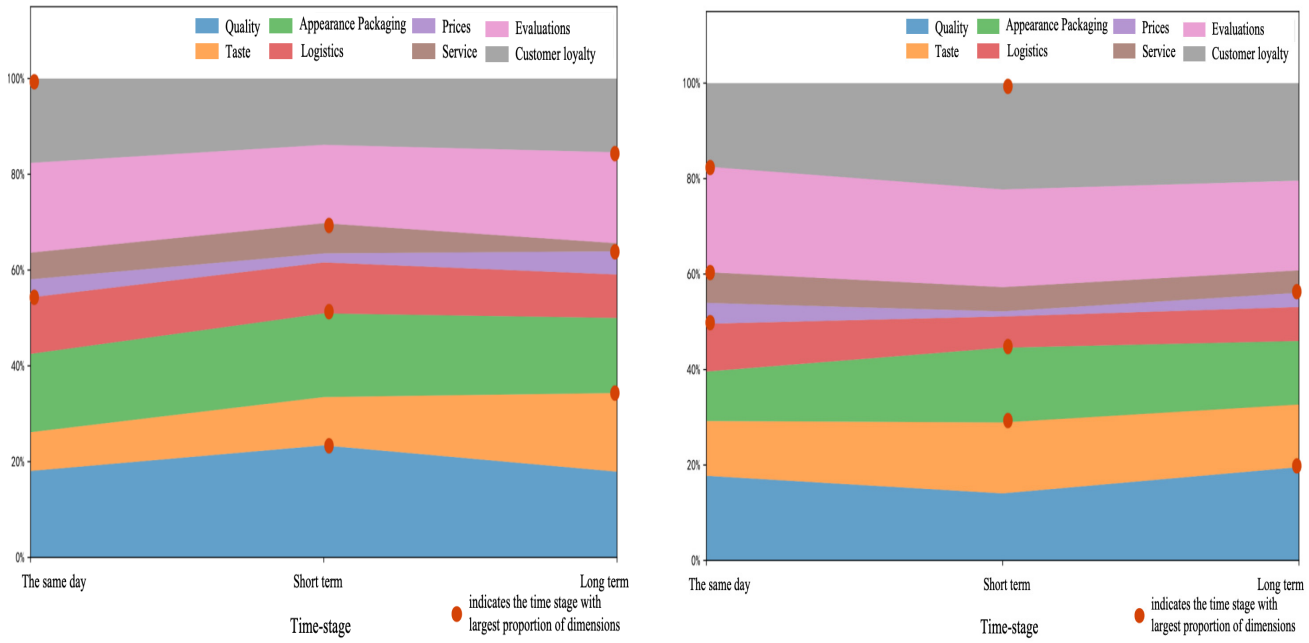


Figure 3. Distribution of feature dimensions of composite reviews in three time stages. (**a**) Initial reviews. (**b**) Additional reviews.

The paper employs a vector space model to illustrate the distribution of customers' interests across different time stages of composite reviews. The five most frequently occurring words are selected as keywords, and their weights are indicated using the TF-IDF index. This approach effectively distinguishes the differences in customers' interest preferences based on their attention to different types and time stages of reviews, as shown in Table 6.

Table 6. Customers' interests in three time stages (top five).

Feature Dimension	The Same Day		Short-Term		Long-Term	
	Initial Reviews	Additional Reviews	Initial Reviews	Additional Reviews	Initial Reviews	Additional Reviews
Quality	(fresh, 0.6351) (quality, 0.3735) (moisture, 0.3650) (stuff, 0.3399) (character, 0.2305)	(fresh, 0.6480) (stuff, 0.5464) (quality, 0.4516) (moisture, 0.4094) (goods, 0.3629)	(fresh, 0.6830) (fresh food, 0.5205) (moisture, 0.4828) (quality, 0.3927) (character, 0.3265)	(moisture,0.4988) (stuff, 0.4571) (fresh, 0.3726) (character, 0.2892) (stuff, 0.2796)	(fresh, 0.6330) (moisture, 0.5656) (quality, 0.4345) (autumn moon, 0.4583) (character, 0.3101)	(goods, 1.0000) (stuff, 0.5000) (fresh, 0.4636) (quality, 0.3910) (moisture, 0.4010)
Taste	(yummy, 0.5660) (texture, 0.3435) (flavour, 0.3323) (very sweet, 0.2160) (tastes, 0.1767)	(yummy, 0.8766) (texture, 0.4157) (flavour, 0.3734) (very sweet, 0.3071) (crunchy-sweet, 0.1823)	(yummy, 0.7232) (flavour, 0.4578) (texture, 0.4590) (sweetness, 0.2150) (very sweet, 0.2079)	(yummy, 1.0000) (texture, 0.4091) (flavour, 0.3183) (not sweet, 0.2399) (unpalatable, 0.2159)	(yummy, 0.9526) (texture, 0.6704) (flavour, 0.4943) (tastes, 0.3060) (icing sugar, 0.3282)	(yummy, 0.8677) (flavour, 0.4811) (texture, 0.4931) (unpalatable, 0.2455) (very sweet, 0.1464)

Table 6. *Cont.*

Feature Dimension	The Same Day		Short-Term		Long-Term	
	Initial Reviews	Additional Reviews	Initial Reviews	Additional Reviews	Initial Reviews	Additional Reviews
Appearance Packaging	(inside, 0.1719) (packaging, 0.5576) (size, 0.2338) (a box, 0.2338) (many, 0.1655)	(packaging, 0.4347) (size, 0.2230) (many, 0.2086) (a box, 0.1835) (very large, 0.1629)	(packaging, 0.6613) (net weight, 0.3752) (single fruit, 0.3237) (size, 0.2806) (100 g, 0.2595)	(packaging, 0.4109) (a box, 0.3072) (open, 0.2815) (inside, 0.2586) (several, 0.2228)	(packaging, 0.5860) (size, 0.3781) (open, 0.2374) (inside, 0.2280) (10 kg, 0.2343)	(a box, 0.2892) (inside, 0.2750) (found, 0.2497) (many, 0.2348) (packaging, 0.2167)
Logistics	(expressage, 0.4276) (received, 0.3604) (logistics, 0.3107) (speed, 0.2665) (delivery, 0.2395)	(expressage, 0.3427) (logistics, 0.3327) (speed, 0.3161) (received, 0.2396) (soon, 0.2054)	(received, 0.5019) (expressage, 0.3892) (logistics, 0.3757) (self-support, 0.3010) (delivered, 0.2359)	(days, 0.2705) (today, 0.2663) (received, 0.2514) (self-support, 0.1584) (delivered, 0.1467)	(received, 0.4641) (logistics, 0.3727) (delivery, 0.2385) (soon, 0.2280) (delivered, 0.2497)	(days, 0.2709) (speed, 0.2384) (received, 0.2324) (dispatch, 0.1917) (delivery, 0.1761)
Prices	(price, 0.2923) (supermarket, 0.2694) (event, 0.1924) (affordable, 0.1396) (cheap, 0.1309)	(price, 0.3712) (supermarket, 0.3359) (affordable, 0.1919) (event, 0.1647) (cheap, 0.1518)	(event, 0.3308) (price, 0.2648) (cost-effective, 0.0907) (value for money, 0.0876) (cheap, 0.0876)	(price, 0.2324) (event, 0.1351) (cheap, 0.0764)	(price, 0.4782) (event, 0.2542) (cheap, 0.2466) (affordable, 0.1840) (value for money, 0.1508)	(price, 0.3266) (value for money, 0.1682) (cheap, 0.1323) (event, 0.1101) (deals, 0.1027)
Service	(store, 0.3100) (customer service, 0.2333) (problem, 0.1782) (after-sales, 0.1733) (service, 0.1593)	(evaluation, 0.2777) (customer service, 0.2847) (after-sales, 0.2000) (store, 0.1951) (problem, 0.1806)	(store, 0.2521) (problem, 0.1766) (customer service, 0.1733) (after-sales, 0.1502) (several, 0.1003)	(customer service, 0.2937) (after-sales, 0.2711) (store, 0.2261) (merchant, 0.1804) (problem, 0.1306)	(store, 0.1535) (attitude, 0.1471) (service, 0.1185) (timely, 0.0963) (customer service, 0.0963)	(store, 0.3503) (customer service, 0.2276) (merchant, 0.1629) (after-sales, 0.1359) (attitude, 0.1057)
Evaluations	(good, 0.7425) (unfilled, 0.6645) (like, 0.3431) (this time, 0.3370) (satisfied, 0.2882)	(good, 0.7318) (like, 0.4078) (everyone, 0.2887) (good, 0.2806) (abysmal, 0.2998)	(unfilled, 0.9415) (good, 0.7845) (like, 0.4051) (satisfied, 0.3320) (good, 0.2802)	(good, 0.6230) (like, 0.2796) (bad, 0.2525) (good, 0.2611) (everyone, 0.2191)	(good, 0.8094) (unfilled, 0.8113) (like, 0.4263) (feel, 0.3910) (worth, 0.3053)	(good, 0.6377) (serviceable, 0.3474) (like, 0.2915) (feel, 0.2978) (so, 0.2875)
Customer loyalty	(evaluations, 0.7541) (buy, 0.4375) (shopping, 0.2471) (once, 0.2336) (before, 0.2338)	(buy, 0.5681) (shopping, 0.3296) (once, 0.3224) (this time, 0.2530) (next time, 0.2199)	(buy, 0.5387) (once, 0.3531) (shopping, 0.3535) (compare, 0.3142) (first, 0.2573)	(buy, 0.5402) (once, 0.3043) (next time, 0.2861) (later, 0.2579) (hope, 0.2335)	(evaluations, 0.8234) (buy, 0.5781) (once, 0.2404) (compare, 0.2335) (shopping, 0.2217)	(buy, 0.5184) (this time, 0.4048) (recommend, 0.3031) (compare, 0.2278) (next time, 0.2018)

4.2. Product Competitiveness Analysis Based on Composite Reviews

The research confidently identifies four regional quadrants: 'Advantage Area', 'Struggle Area', 'Opportunity Area', and 'Waiting Area' (Figure 4). These quadrants characterise product competitiveness based on online reviews and feature dimensions. The 'Advantage Area' is the feature dimension that customers frequently discuss and pay the most attention to. It is the most effective in promoting customers' perceived value and determines the core competitive advantage of the business. Conversely, 'Struggle Areas' represent feature dimensions that lack information but receive considerable attention. These feature dimensions provide valuable product information, and businesses may therefore seek to direct their customers to post adequate and high-quality reviews in order to bring them closer to the 'Advantage Areas'. The 'Opportunity Area' is a dimension that indicates that customers are less enthusiastic in their discussions and pay less attention to the features. Consequently, merchants must adjust their resources in real time according to the market dynamics. 'Waiting Area' is a feature dimension that is frequently discussed but paid less attention to, and it is an important key direction for merchants to improve service competition and seize market opportunities in the future.

The feature dimensions in the initial reviews, such as 'appearance packaging' and 'logistics', are more competitive than those in the additional reviews. This may be attributed to the fact that customers have not yet had the opportunity to fully appreciate the product, with only a superficial understanding of its packaging and logistics features. These factors can initially influence the customer experience. The initial feedback indicated that the 'evaluations' aspect was highly favoured, while the 'quality' aspect was the most

concerning. Conversely, the additional feedback was most positive about the 'taste' aspect, but expressed the most concern about the 'evaluations' aspect. An additional review is a characterisation of the customer's experience after using the product for a period of time. This allows the customer to assess the product in terms of its quality and taste, which reflects their original intention and need to purchase the product.

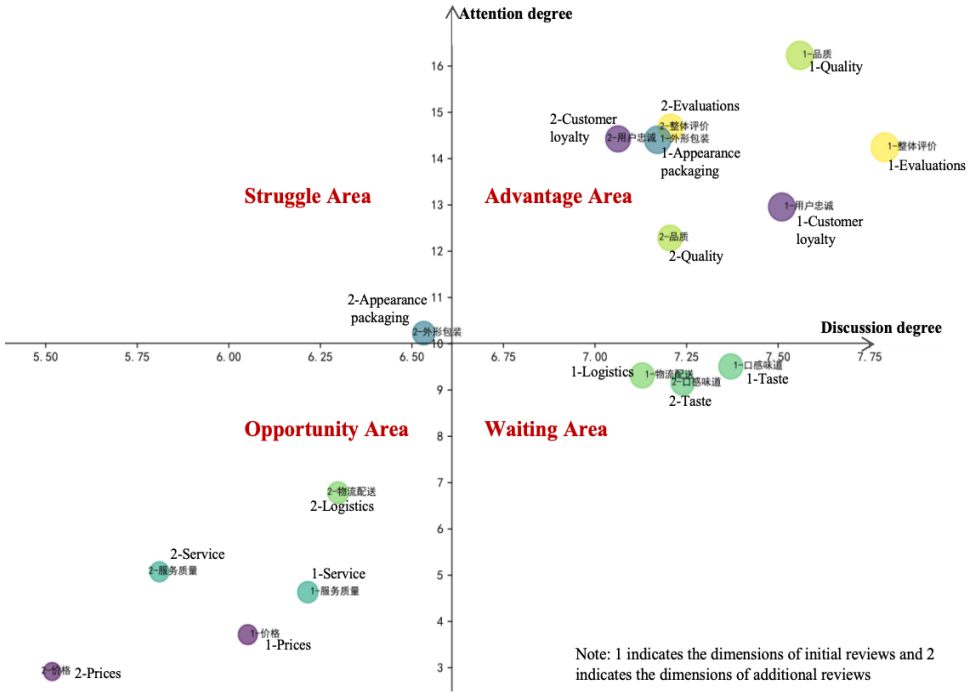


Figure 4. Competitiveness analysis of feature dimensions.

4.3. Sentiment Analysis Based on Composite Reviews

Online reviews can significantly impact customer preferences and purchase intentions [7]. This research analyses sentiment value differences in three types of reviews: same-day, short-term, and long-term for initial and additional reviews, respectively. And then, the sentiment mean value was calculated by averaging the sentiment scores of the three stages. The research findings indicate that emotional attitudes in both initial and additional reviews became increasingly negative over time. Despite the fact that subsequent reviews were more diverse, they still carried negative emotions, resulting in an overall low emotional mean (Figure 5). This emphasises the importance for merchants to prioritise the long-term interests of their customers and to make efforts to enhance the quality, appearance packaging, and taste of their services to obtain positive evaluations. This shift in sentiment, in addition to reflecting the merchant's shortcomings, may also be correlated with a change in the customer's perception of the product. As the customer becomes more aware of the product over time, this has a direct impact on the customer's satisfaction and the sentiment expressed in the review.

Additionally, the sentiment mean value for additional reviews was significantly lower when compared to the level of initial reviews. Additional reviews are usually additional evaluations of the product experience. The sentiment value of the additional reviews is generally lower than that of the initial reviews, indicating that the product does not fully align with the customer's needs. Negative reviews that appear in both the initial and additional reviews will undoubtedly reduce potential customers' confidence in shopping. Therefore, businesses should maintain close service communication with customers to ensure that they receive an effective solution to their problem after providing a negative evaluations. If there is an initial review of the negative emotions, the business should be prompt in appeasing the customer, to prevent the deterioration of emotions. Customers

who receive a satisfactory service are more likely to provide positive feedback, which can be further incentivised through offers such as coupons. Furthermore, fluctuations in the types of fresh fruits due to seasonal factors may result in seasonal patterns in services related to fresh fruits. The resulting seasonal fluctuations in prices may also lead to an increase in customers' price sensitivity, which in turn affects their review sentiment. The research presented in this paper identifies the potential avenues for improvement in the service provided by current fresh food platforms, as customers' emotional attachments to fresh fruits tends to decrease over time. Nevertheless, in the case of fresh food platforms that consistently deliver high-quality products, the emotional intensity of customers does not necessarily decline gradually. Rather, it remains at a high level.

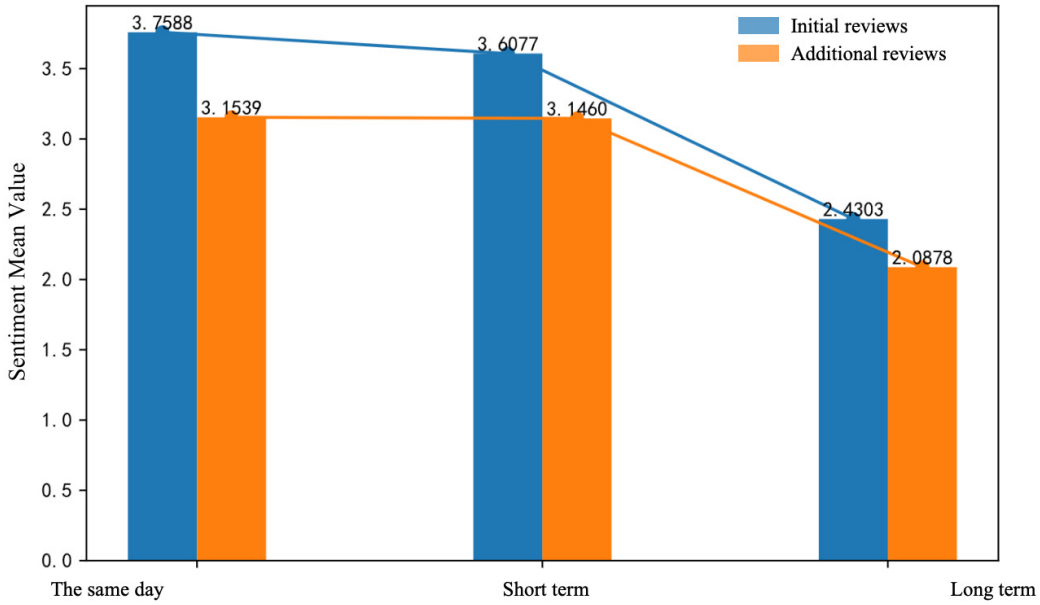


Figure 5. Comparison of emotional means in three time stages.

5. Conclusions

This research employs text mining and sentiment analysis of online reviews of fresh produce on e-commerce platforms to identify shifts in customer preferences and differences in composite reviews at different times. The findings reveal that customers' interests in purchasing fresh produce online are evolving. Merchants should direct their attention to enhancing the eight quality dimensions that customers primarily consider when making online purchases of fresh products. These dimensions include quality, taste, appearance packaging, logistics, prices, service, evaluations, and customer loyalty. This will result in an enhancement of the products in question.

At the same time, assessing the competitiveness of product features in different time stages based on the degree of attention and discussion intensity can assist businesses in adjusting their resources according to customer interests, developing corresponding marketing strategies, and providing accurate and effective information services to their customers. It can be observed that customer preferences exhibit heterogeneity in various types of feedback and at different stages. Furthermore, the level of attention paid by customers to service features and the intensity of discussion may change due to shifts in customer preferences and interests.

It is possible that the customer's emotional attitude may become increasingly negative over time, which could result in a greater diversity of reviews with a negative sentiment bias. This suggests that the observed change in customer satisfaction with online fresh fruit consumption can be attributed to perceived differences in the consumption experience across various characteristic dimensions. Merchants must maintain close communication with customers in order to ensure effective problem resolution and to prevent negative eval-

uations. If a customer expresses negative emotions in their first review, the merchant must promptly appease them to maintain positive customer relations and improve satisfaction.

Furthermore, this research has identified that product competitiveness can be analysed based on initial and additional reviews, which can be categorised into four quadrants: 'Advantage Area', 'Struggle Area', 'Opportunity Area', and 'Waiting Area'. The feature dimensions in the 'Advantage Area' serve as the primary indicators of the core competitiveness of the merchant's products. In contrast, the feature dimensions in the 'Struggle Area' require customers to post sufficient and high-quality reviews in order to enhance the core competitiveness of the products. The 'Opportunity Area' represents product dimensions that are less frequently reviewed and discussed. In contrast, the 'Waiting Area' is a feature dimension that is frequently discussed but receives less attention. It is a priority for companies to improve their service competitiveness and seize market opportunities in the future. This research proposes a product competitiveness quadrant that can serve as a framework for businesses seeking to enhance their products and operations with confidence.

The present research examined the temporal evolution of sentiment means values. Subsequent studies should combine the validity and sentiment scores of each dimension in order to refine the analysis. Furthermore, future studies should explore the influence of demographic characteristics and product presentation on customer preferences.

Author Contributions: Conceptualization, H.L.; methodology, Z.W. and Z.L.; software, Y.C.; writing—original draft, Z.W. and Y.C.; writing—review and editing, Z.W. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Education Science Planning Youth Project of the Ministry of Education (Grant No.: EIA210424).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hu, N.; Liu, L.; Zhang, J.J. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Inf. Technol. Manag.* **2008**, *9*, 201–214. [CrossRef]
2. Chen, Y.; Xie, J. Third-party product review and firm marketing strategy. *Mark. Sci.* **2005**, *24*, 218–240. [CrossRef]
3. Lin, Y.; Jen, L. The impact of consumer heterogeneity in the product life cycle on the diffusion patterns of user reviews and sales. *J. Retail. Consum. Serv.* **2024**, *76*, 103558.
4. Li, Y.; Wang, K.; Xu, F.; Fan, C. Management of trade-in modes by recycling platforms based on consumer heterogeneity. *Transp. Res. Part E Logist. Transp. Rev.* **2022**, *162*, 102721. [CrossRef]
5. Ho-Dac, N.N.; Carson, S.J.; Moore, W.L.; Ho-Dac, S.J.C.N.N.; Nam, H.; Kannan, P. The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *J. Mark.* **2013**, *77*, 37–53. [CrossRef]
6. Liu, Y.; Gan, W.X.; Zhang, Q. Decision-making mechanism of online retailer based on additional online comments of consumers. *J. Retail. Consum. Serv.* **2021**, *59*, 102389. [CrossRef]
7. Wang, S.; Ye, S.; Dong, Y.; Wang, H. Study on the Impact of Additional Comments on Consumers' Purchase Decisions: Comparative Analysis Based on Hedonic and Practical Products. In Proceedings of the 2019 2nd International Conference on E-Business, Information Management and Computer Science, Kuala Lumpur, Malaysia, 12 August–12 October 2019; pp. 1–6.
8. Zhao, Y.; Wen, L.; Feng, X.; Li, R.; Lin, X. How managerial responses to online reviews affect customer satisfaction: An empirical study based on additional reviews. *J. Retail. Consum. Serv.* **2020**, *57*, 102205. [CrossRef]
9. Chen, C.; Zhang, D.; Zhang, F.; Zhang, F. The interaction effect of online composite reviews and time pressure on information persuasiveness: A heuristic-systematic perspective. *Electron. Commer. Res. Appl.* **2023**, *59*, 101269. [CrossRef]
10. Shi, W.; Gong, X.; Zhang, Q.; Wang, L. A comparative study on the first-time online reviews and appended online reviews. *J. Manag. Sci.* **2016**, *29*, 45–58.
11. Wang, C.; He, S.; Wang, K. Research on how additional review affects perceived usefulness of review. *J. Manag. Sci.* **2015**, *28*, 102–114.
12. Zhang, D.; Ma, J.; He, Z. The effect of emotional changes in composite reviews on consumers' information adoption from a dual perspective. *Ind. Manag. Data Syst.* **2024**, *124*, 2205–2229. [CrossRef]

13. Zhou, H.; Li, S. The impact of online additional comments on consumers' information adoption. *Sociol. Mind* **2017**, *7*, 60–71. [CrossRef]
14. Luo, Y.; Li, Y.; Ye, Q. Impacts of online additional reviews on the sales volume of cross-border pharmaceutical e-commerce platforms. *J. Glob. Inf. Technol. Manag.* **2022**, *25*, 83–101. [CrossRef]
15. Hossain, M.S.; Rahman, M.F. Customer sentiment analysis and prediction of insurance products' reviews using machine learning approaches. *FIIB Bus. Rev.* **2023**, *12*, 386–402. [CrossRef]
16. Chang, V.; Liu, L.; Xu, Q.; Li, T.; Hsu, C.H. An improved model for sentiment analysis on luxury hotel review. *Expert Syst.* **2023**, *40*, e12580. [CrossRef]
17. Li, H.; Bruce, X.B.; Li, G.; Gao, H. Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews. *Tour. Manag.* **2023**, *96*, 104707. [CrossRef]
18. Purnawirawan, N.; De Pelsmacker, P.; Dens, N. Balance and sequence in online reviews: How perceived usefulness affects attitudes and intentions. *J. Interact. Mark.* **2012**, *26*, 244–255. [CrossRef]
19. Gao, M.; Liu, K.; Wu, Z. Personalisation in web computing and informatics: Theories, techniques, applications, and future research. *Inf. Syst. Front.* **2010**, *12*, 607–629. [CrossRef]
20. Savci, P.; Das, B. Prediction of the customers' interests using sentiment analysis in e-commerce data for comparison of Arabic, English, and Turkish languages. *J. King Saud Univ. Comput. Inf. Sci.* **2023**, *35*, 227–237. [CrossRef]
21. Wijaya, I.W.R. Development of conceptual model to increase customer interest using recommendation system in e-commerce. *Procedia Comput. Sci.* **2022**, *197*, 727–733. [CrossRef]
22. Liu, Z.; Chen, X.; Sun, M. Mining the interests of Chinese microbloggers via keyword extraction. *Front. Comput. Sci.* **2012**, *6*, 76–87. [CrossRef]
23. Cena, F.; Likavec, S.; Osborne, F. Propagating user interests in ontology-based user model. In *AI*IA 2011: Artificial Intelligence Around Man and Beyond, Proceedings of the XIIth International Conference of the Italian Association for Artificial Intelligence, Palermo, Italy, 15–17 September 2011*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 299–311.
24. Kim, Y.; Shim, K. TWILITE: A recommendation system for Twitter using a probabilistic model based on latent Dirichlet allocation. *Inf. Syst.* **2014**, *42*, 59–77. [CrossRef]
25. Gen, L.; Jin, G.; Li, Y.; Sun, W.; Ma, S. Research on Personalized Recommendation of University Library Literature Resources Based on Improved Content-based Filtering Algorithm. *Libr. Inf. Serv.* **2018**, *62*, 112.
26. Shi, C.; Hu, B.; Zhao, W.X.; Philip, S.Y. Heterogeneous information network embedding for recommendation. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 357–370. [CrossRef]
27. Zhang, P.; Zhang, Z.; Tian, T.; Wang, Y. Collaborative filtering recommendation algorithm integrating time windows and rating predictions. *Appl. Intell.* **2019**, *49*, 3146–3157. [CrossRef]
28. Karthik, R.V.; Ganapathy, S. A fuzzy recommendation system for predicting the customers interests using sentiment analysis and ontology in e-commerce. *Appl. Soft Comput.* **2021**, *108*, 107396. [CrossRef]
29. Carmel, Y.; Patt-Shamir, B. Comparison-based interactive collaborative filtering. In *Structural Information and Communication Complexity, Proceedings of the 22nd International Colloquium, Montserrat, Spain, 14–16 July 2015*; Springer: Montserrat, Spain, 2015; pp. 429–443.
30. Tang, S.; Gong, L.; Xie, J.; Liu, Y. A competitiveness evaluation method of product based on technology maturity and development trend. In Proceedings of the 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), Siem Reap, Cambodia, 18–20 June 2017; pp. 675–680.
31. Liu, Z.; Qin, C.X.; Zhang, Y.J. Mining product competitiveness by fusing multisource online information. *Decis. Support Syst.* **2021**, *143*, 113477. [CrossRef]
32. Xu, K.; Liao, S.S.; Li, J.; Song, Y. Mining comparative opinions from customer reviews for competitive intelligence. *Decis. Support Syst.* **2011**, *50*, 743–754. [CrossRef]
33. Wu, J.; Zhang, J.; Zhao, N. How to boost e-commerce for poverty alleviation? A perspective on competitiveness analysis using online reviews. *Electron. Commer. Res.* **2023**, 1–32. [CrossRef]
34. Kim, Y.; Dwivedi, R.; Zhang, J.; Jeong, S.R. Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5. *Online Inf. Rev.* **2016**, *40*, 42–61. [CrossRef]
35. Lu, X.; Guo, C.; Liu, J. Study on Identification of Enterprise Product Level Competitor Based on Sentiment Analysis: Taking User Reviews for Data Resources. *Libr. Inf. Serv.* **2016**, *60*, 83–90.
36. Xu, X.; Wang, X.; Li, Y.; Haghighi, M. Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *Int. J. Inf. Manag.* **2017**, *37*, 673–683. [CrossRef]
37. Syakur, M.A.; Khotimah, B.K.; Rochman, E.M.S.; Satoto, B.D. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Surabaya, Indonesia, 2018; Volume 336, p. 012017.
38. Umargono, E.; Suseno, J.E.; Gunawan, S.V. K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula. In *The 2nd International Seminar on Science and Technology (ISSTEC 2019)*; Atlantis Press: Yogyakarta, Indonesia, 2020; pp. 121–129.

39. Nainggolan, R.; Perangin-angin, R.; Simarmata, E.; Tarigan, A.F. Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method. In *Journal of Physics: Conference Series*; IOP Publishing: Medan, Indonesia, 2019; Volume 1361, p. 012015.
40. Lu, A.; Lin, W. An integrated analysis of topical and emotional evolution of microblog public opinions on public emergencies. *Libr. Inf. Serv.* **2017**, *61*, 120.
41. Xu, L.; Lin, H.; Pan, Y.; Ren, H.; Chen, J. Constructing the affective lexicon ontology. *J. China Soc. Sci. Tech. Inf.* **2008**, *27*, 180–185.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.