ASME Journal of Mechanical Design
Online journal at:
https://asmedigitalcollection.asme.org/mechanicaldesign

Analysis of Brand Effects in Data-Driven Design Based on Online Reviews

Seyoung Park
Enterprise Systems Optimization Laboratory,
Department of Industrial and Enterprise Systems Engineering,
University of Illinois Urbana-Champaign,
Urbana, IL 61801
e-mail: seyoung7@illinois.edu

Harrison Kim[1]
Enterprise Systems Optimization Laboratory,
Department of Industrial and Enterprise Systems Engineering,
University of Illinois Urbana-Champaign,
Urbana, IL 61801
e-mail: hmkim@illinois.edu

Recently, online user-generated data have emerged as a valuable source for consumer product research. However, most studies have neglected the brand effect, although it is a significant factor in conventional market research. This paper demonstrates the importance of brands in data-driven design using online reviews. Specifically, this study utilizes game theory and suggests a game setting representing market competition. Elements of the game are determined based on online data analysis. The proposed approach consists of four stages. The first stage divides online customers into different segments and analyzes them to extract the feature importance of each brand in each segment. The importance is based on the positive term frequency of features, and it becomes the customer's partial utility for each feature. The second stage defines the specification of product candidates and calculates their costs. This study refers to real market datasets (bill of materials) available online. At this point, the game is all set. The third stage finds the Nash equilibrium of the designed game, and the final stage compares the optimal strategy for a product portfolio with and without brand consideration. The suggested approach was tested on smartphone reviews from Amazon. The result shows that the lack of brand consideration leads a company to choose a non-optimal product strategy, illustrating the significance of the brand factor. [DOI: 10.1115/1.4063288]

Keywords: data-driven design, online review, brand effect

1 Introduction

A brand is a significant factor in today's marketplace. The company's success is greatly dependent on the power of its brand [1], which is measured by various indices such as brand awareness, brand image, brand preference, and brand relevance [2]. Na et al. [3] explained a brand power model as a combination of brand awareness and brand image. Aaker [2] pointed out that the most common basis of market competition is to win the brand preference battle [2]. A renowned American business magazine, Forbes, publishes the brand values of global companies every year [4]. The American Customer Satisfaction Index, an influential survey published by the University of Michigan, reports satisfaction benchmarks for brands in various industry sectors [5].

In industry, companies consider their brand indices when developing new products. They usually hire a market research firm [6] and conduct surveys to obtain data, which quantifies the market player's brand power in various aspects. The company recognizes its market position based on the result and devises the strategy for its products. For example, in the automobile market, Toyota has strengths in the affordable price and design that appeals to young customers. On the other hand, Mercedes has strengths in that the brand has a touch of class, and its design appeals to older people [7]. Therefore, these two brands will target different customer segments and setup different strategies to maximize the value of their products for the target customers.

Recently, many studies have been utilizing online user-generated data in their research. In data-driven design [8], these studies analyze online data to understand customers' preferences and draw design implications. The resultant implications include feature importance [9,10], usage [11], spec guidance [12,13], and ideas for new features [14,15]. However, the previous studies did not consider brand influence while it is a significant factor in the industry. They analyzed the whole product base assuming that there is no difference between brands, i.e., different brands have similar strengths and weaknesses. It is a significant gap between the industry and research.

This paper aims to investigate the effect of a brand factor in product design based on user-generated data. The proposed approach mimics real-world market competition by setting up a game, which includes customer utility and product portfolio. This study extracts customer utility by analyzing online reviews and determines product spec configuration and costs based on resources available online. A case study with smartphone data shows that the reflection of the brand factor alters the optimal strategy for different brands.

In Sec. 2, relevant studies will be introduced and reviewed. Section 3 will explain the details of the proposed methodology. In Sec. 4, the methodology will be tested on real-world datasets. Section 5 will compare the resultant product portfolio with and

[1]Corresponding author.
Contributed by the Design Automation Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received May 3, 2023; final manuscript received August 20, 2023; published online October 6, 2023. Assoc. Editor: Christopher Mccomb.

without the brand factor and discuss the influence of the brand. Finally, in Sec. 6, the contribution of this research will be summarized, and future works will be discussed.

2 Literature Review

In this section, research in data-driven design based on user-generated data will be presented, and their limitations will be discussed.

2.1 Data-Driven Design. Data-driven design means that the design is based on the use of data science algorithms supporting specific phases of the product development process [16]. Product designers can harness their organization's competitive edge by uncovering patterns and novel insights from huge and highly contextualized data [8]. Among various types of data sources in this area, online user-generated data have become a popular resource for consumer product design. Many studies utilize online data because of its strength in time and cost-efficiency compared to conventional methods such as surveys and interviews. There exist various approaches to extracting design implications from online data.

Chaklader and Parkinson [12] proposed a methodology to extract proper size specifications for headset products from online reviews. They selected reviews with positive sentiments for the product size and calculated the average rating of these reviews. The authors suggested proper specs by comparing this value with the average rating of total reviews.

Some studies focused on the different importance of product features. Suryadi and Kim [9] analyzed online reviews to understand the influence of product features on product sales ranking. First, they identified product features in data using Word2Vec [17] and clustering. Then, the authors quantified each reviewer's sentiment for the identified features. Each review has a set of {feature: sentiment score} pairs. The scores became input data to linear regression, and the output data are sales ranking. In the regression result, the coefficients indicate the effect of each product feature on sales ranking. Wang and Chen [18] studied the effect of product features on customers' purchase behavior. The authors generated choice sets using online user data and constructed multinomial logit models [19]. The coefficients in the result show the influence of product attributes on the customer's purchase decision. Joung and Kim [10] suggested a methodology to identify the importance of product features based on review ratings. They extracted feature keywords from online reviews using latent Dirichlet allocation [20] and analyzed each customer's sentiment for each feature. Then, the authors built a neural network model, where the input data are the sentiment scores, and the output data are the customer's rating for the product. By interpreting the trained model with SHapley Additive exPlanations [21,22], the authors obtained influence scores of product features on the review ratings.

Another approach is to discover new product features from user-generated data. Tuarob and Tucker [14] suggested a methodology to extract ideas for new features from social media data. They extracted ground-truth features from product spec documents and user-discussed features from Twitter data. Then, the authors identified latent features and detected lead users on Twitter based on these features. The suggested methodology discovered new smartphone features by analyzing the lead users' Twitter mentions. Goldberg and Abrahams [15] presented a method that sources product innovation ideas from online reviews. They adopted and revised the attribute mapping framework. It differentiates product attributes based on customer sentiments (positive/negative/neutral) and attribute types (basic/discriminators/energizers). The reviews mentioning product features were analyzed and assigned to one of the categories in the framework. The result suggested candidates for new product features and their priorities.

As shown previously, the previous studies proposed various methods for extracting design implications from online user-generated data. However, they have a limitation since they disregarded a brand factor assuming the same characteristics for all products. This study incorporates the brand factor when analyzing the user data and shows the difference in product strategies made by brands.

2.2 Brands in User-Generated Data. It can be questioned whether the brand effect is reflected in user-generated data. Regarding this question, some studies showed the existence of brand effects on user data by analyzing online sources. Jin et al. [23] analyzed Amazon reviews and extracted customer sentiments for mobile phone attributes. They compared products of different brands and analyzed whether one product is more favorable than a competitive one at the feature level. The result shows that each brand has different strengths and weaknesses in terms of product features. Tuarob and Tucker [24] conducted similar research using Twitter data. They listed the top 10 strong/weak features of each manufacturer based on customer satisfaction. The result shows that different brands have different lists. Nuortimo and Harkonen [25] suggested a method that extracts a brand index from text data on social media. The authors analyzed the sentiment in user opinions mentioning the target brand and computed the percentual share of negative opinions. They showed that different brands have different percentage values, and a lower percentage means a higher brand index. The resultant brand index was validated by the consistency with the Forbes brand index [4]. Alzate et al. [26] proposed a method that extracts brand image and brand positioning from online reviews. They analyzed the text data using the lexicon-based linguistic inquiry and word count program and clustered brands in the cosmetic industry. The resultant brand positioning map illustrates differences in four brand clusters.

Although the above studies showed that the brand factor is reflected in user-generated data, the influence of the brand in industrial applications has been rarely discussed. This study investigates how the brand factor affects product design or a company's strategy. Specifically, this paper compares the company's product portfolio with and without brand consideration and demonstrates the importance of brands in data-driven design.

3 Methodology

This study analyzes the influence of the brand in data-driven design, especially using online user-generated data. The game theory approach is adopted to compare the company's product portfolio with and without brand consideration. This section provides detailed explanations for game settings and the four stages of the proposed methodology. The first stage calculates the customer's partial utility, i.e., the utility for each product feature, by brand. The second stage defines product candidates for each brand and calculates customer utility for them. The next stage analyzes the competition between brands and finds the optimal product portfolio strategy for both brands. The final stage compares the optimal portfolio with and without brand consideration.

Regarding the market competition, the game setting in Ref. [27] is modified and used in this study. Table 1 shows the framework of the game setting. Each brand has multiple product candidates for its portfolio. The cost of each product is estimated to calculate a customer's utility per unit price. The customer base is segmented into different groups. It is assumed that customers in different segments have different utilities for the same product. The customer utility is calculated based on Eq. (1), where b, i, j, and k represent the brand, product, customer segment, and product feature, respectively. w_{jk}^b is the weight (importance) that segment j has for product feature k in brand b. x_{ik}^b represents the spec value of feature k of product i in brand b.

$$U_{ij}^b = \sum_{k=1}^{K} w_{jk}^b \times x_{ik}^b \quad (1)$$

Table 1 Game settings

Player	Product	Cost	Customer utility S_1	S_2	S_3	$\cdots$	S_j
Brand (b)	P_1^b	C_1^b	U_{11}^b	U_{12}^b	U_{13}^b	$\cdots$	U_{1j}^b
	P_2^b	C_2^b	U_{21}^b	U_{22}^b	U_{23}^b	$\cdots$	U_{2j}^b
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
	P_i^b	C_i^b	U_{i1}^b	U_{i2}^b	U_{i3}^b	$\cdots$	U_{ij}^b

Note: P_i^b = product candidate i for brand b, S_j = customer segment j, and U_{ij}^b = the utility of segment j for P_i^b calculated by Eq. (1).

Filling out the utility column in Table 1 requires three types of datasets: (i) segmented customers; (ii) feature weights by brand in each segment; and (iii) product spec configurations. Among them, (i) and (ii) belong to customer analysis, and (iii) is determined in product analysis.

3.1 Customer Analysis. Figure 1 shows the process of customer analysis, which consists of three stages. In the first stage, product features of customer interests are extracted from online user-generated data. Next, customers' sentiments for these features are analyzed, and customers with similar interests are grouped. In the final stage, feature weights for each brand are calculated based on the previously obtained sentiment analysis. The first and second stages are based on the authors' previous works. It is considered necessary to summarize the method while the details are available in Refs. [28,29].

The first stage is feature extraction based on phrase embedding and clustering [28]. Initially, the words in the online review data are embedded into vectors by Word2Vec [17]. Next, the method extracts phrases from the review data. Only bi-grams are considered in this study, as in Ref. [28]. The result contains all bi-grams, including noun-noun, adjective-noun, etc. Then, the method filters noise phrases using product manuals. The phrases with the word never mentioned in product manuals are removed. The remaining phrases are embedded into a vector space by Eq. (2), a weighted sum of word vectors ($\vec{W}_i$ and $\vec{W}_j$). The weight α is calculated based on the frequency of two words in the manual documents. Finally, the phrase vectors are grouped into clusters. The phrases in the feature-related clusters become feature keywords and will be used in the next stage.

$$
\begin{aligned}
\text{Phrase} &= \alpha_i \times \vec{W}_i + \alpha_j \times \vec{W}_j \\
\alpha_i &= \frac{\text{Freq}(W_i)}{\text{Freq}(W_i) + \text{Freq}(W_j)}
\end{aligned}
\tag{2}
$$

The second stage is market segmentation. Among two approaches—customer segmentation [30] and product segmentation [31]—this paper adopts the prior one based on network analysis [29]. The method in Ref. [29] consists of three steps. The first step analyzes each customer's sentiment for product features by detecting review sentences with feature keywords and then measuring the sentiment scores of those sentences using VADER [32]. The resulting data are a list of sentiment scores, and it is converted into a sentiment polarity vector. Figure 2 illustrates an example of this process. Let us assume that the target features are [screen, camera, price]. The presented review mentions the camera and price, and the corresponding sentiments are 0.48 and 0.64, respectively. So, the initial result is [0, 0.48, 0.64]. Then it is converted into [0, 0, 1, 0, 1, 0] based on the feature and polarity. For example, C + is 1 because the customer expresses a positive opinion about the camera. P − is 0 because there is no negative opinion about the price. Since the review does not mention the screen feature, both S + and S − are 0. The obtained vector becomes the customer attribute. The customers with all-zero attributes are removed from the data. In the next step, the method constructs a network by connecting customers with similar attributes. Reference [29] presented a formula in Eq. (3) to connect customers. $\text{Sim}^i(i, j)$ is the similarity score between customers i and j from customer i's perspective. It calculates the ratio of the number of topics (feature + sentiment) common in two customers to the number of topics mentioned by customer i. The second line of Eq. (3) is the mathematical expression of this concept, where a_k^i denotes the attribute value of customer i for topic k. If the similarity score is greater than or equal to a threshold value α for both customers, then two nodes corresponding to these customers are connected by an edge. Otherwise, the nodes cannot be connected.

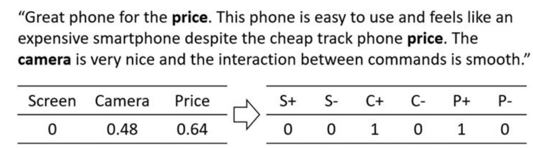

Fig. 2 Sentiment analysis process

$$
\begin{aligned}
\text{Sim}^i(i, j) &= \frac{\#\text{topics common in customer } i, j}{\#\text{topics mentioned by customer } i} \\
\text{Sim}^i &= \frac{\sum_{k=1}^{n} a_k^i a_k^j}{\sum_{k=1}^{n} a_k^i}, \quad \text{Sim}^j = \frac{\sum_{k=1}^{n} a_k^i a_k^j}{\sum_{k=1}^{n} a_k^j} \\
\text{E}_{ij} &= \begin{cases} 1 & \text{if } \text{Sim}^i \geq \alpha \text{ and } \text{Sim}^j \geq \alpha \\ 0 & \text{else} \end{cases}
\end{aligned}
\tag{3}
$$

Figure 3 illustrates the process of network construction. Customers 1 and 2 have scores of 0.67 (2/3) and 0.50 (2/4), respectively. If we set $\alpha = 0.5$, they are connected because both customers have a similarity score greater than or equal to 0.5. Customers 2 and 3 cannot be connected since customer 2's score is less than 0.5. Customers 1 and 3 are connected because their scores are greater than 0.5. In the third step, the method conducts network clustering to divide a customer base into several groups with similar characteristics. The optimal number of segments is automatically determined by modularity clustering [33,34].

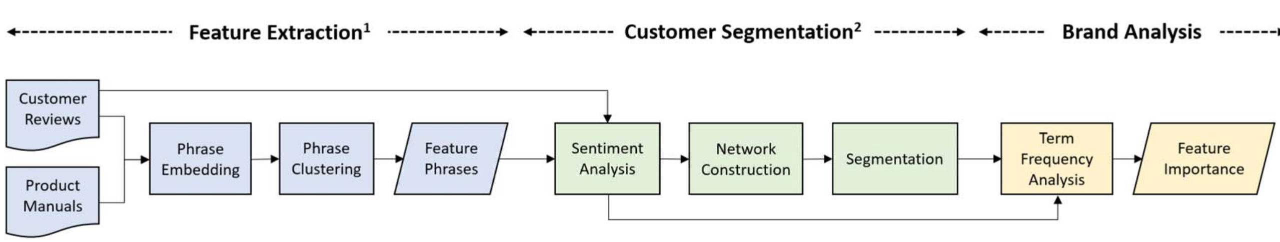

Fig. 1 Flowchart of customer analysis

The final stage is brand analysis, where feature weights for each brand are analyzed. This research aims to demonstrate the influence of brands rather than develop a new method for analyzing feature weights for different brands. Therefore, this study uses the term frequency (TF), a simple method for importance evaluation [35]. Specifically, the number of positive reviews for each feature is counted and normalized, as shown in Eq. (4). TF^b_{jk} denotes the number of reviews expressing positive sentiments for a specific feature k of products in brand b among customers in segment j. In the same manner, w^b_{jk} is the feature weight that customers in segment j have for a specific feature k of products in brand b. The TF for a certain feature is zero in some segments. Therefore, an offset of 1 is applied to all TF. The final value is the normalized TF for each feature with offset.

$$w^b_{jk} = \frac{(\mathrm{TF}^b_{jk} + 1)}{\sum_{k'=1}^{K} (\mathrm{TF}^b_{jk'} + 1)} \tag{4}$$

3.2 Product Analysis. The remaining part of the game setting is product data. Table 1 requires spec configurations for product candidates (P^b_i) and their costs (C^b_i). In the industry, available spec options for each feature are determined by the internal sourcing department of a company. The cost of each component is dependent on the estimated sales volume. Therefore, the data for product candidates may be different by brand (company). In this study, available spec options are defined based on released products in the market. The cost of each spec is obtained from the bill of materials (BoM) available online. This paper suggests three scenarios for product candidates.

- Case 1: Two brands have the same product candidates.
- Case 2: Two brands have the same product candidates with different spec configurations from case 1.
- Case 3: Two brands have different product candidates.

Because this study aims to show the importance of the brand factor, we want to emphasize the changes made by the brand aside from other factors. Therefore, cases 1 and 2 assume the same product candidates to show how differences in customers' perception of two brands affect the company's optimal strategy even when two brands have the same condition. Once spec configurations are determined, the utility of each product candidate can be calculated based on Eq. (5), where U^b_{ij} is the segment j's utility for product candidate i of brand b. Basically, the utility is a weighted sum of spec values. The first term is the weight obtained from the result of customer sentiment analysis (Eq. (4)). The second term x^b_{ik} represents the spec value of feature k of product i in brand b.

$$U^b_{ij} = \sum_{k=1}^{K} \left(\frac{(\mathrm{TF}^b_{jk} + 1)}{\sum_{k'=1}^{K} (\mathrm{TF}^b_{jk'} + 1)} \times x^b_{ik} \right) \tag{5}$$

3.3 Portfolio Analysis. A game theory approach [36–38] is an appropriate way to mimic the competition of different companies. This section presents a game theory-based method for portfolio analysis, which consists of three steps: (i) define feasible strategies, (ii) construct a payoff matrix, and (iii) find Nash equilibrium (NE).

First, feasible strategies should be defined. In industry, companies have limitations in determining product portfolios due to restrictions, such as fixed budgets and component sourcing failures. Therefore, they need to determine available strategies under the given situation. For example, let us assume that brand 1 has four product candidates ($P^1_1, P^1_2, P^1_3, P^1_4$) and is allowed to release up to two new products due to the limited budget. In this case, the feasible strategy is a single candidate or a combination of two products out of four candidates, i.e., [$P^1_1, P^1_2, P^1_3, P^1_4, (P^1_1, P^1_2), (P^1_1, P^1_3), (P^1_1, P^1_4), (P^1_2, P^1_3), (P^1_2, P^1_4), (P^1_3, P^1_4)$].

Next, the method constructs a payoff matrix, which shows the benefit of each brand when two brands choose certain strategies. Table 2 shows an example of a payoff matrix. Brand 1 is a row player. The first row represents the game where brand 1 chooses strategy 1 (Z^1_1), and the second row is the situation when brand 1 selects strategy 2 (Z^1_2). Brand 2 is a column player. The first and second columns can be interpreted in the same manner. The value in each cell is the payoff for each brand. When brands 1 and 2 choose strategies Z^1_1 and Z^2_1, respectively, the payoff for brand 1 is 0.2, and that for brand 2 is 0.4. Regarding payoff calculation, Sadeghi and Zandieh [27] presented a function, shown in Eq. (6), for product portfolio management. This study employs this function because we focus on the optimal strategy for product portfolios. Also, changes in customer utility by brand can be reflected in the payoff calculation.

$$\begin{aligned} f_1(Z^1_x, Z^2_y) &= \sum_{j=1}^{J} \sum_{i=1}^{I^1} \left(\frac{U^1_{ij}}{C^1_i} \times \frac{e^{\mu U^1_{ij}}}{\sum_{c=1}^{I_{\mathrm{com}}} e^{\mu U_{cj}}} \times Q_j \right) \\ f_2(Z^1_x, Z^2_y) &= \sum_{j=1}^{J} \sum_{i=1}^{I^2} \left(\frac{U^2_{ij}}{C^2_i} \times \frac{e^{\mu U^2_{ij}}}{\sum_{c=1}^{I_{\mathrm{com}}} e^{\mu U_{cj}}} \times Q_j \right) \end{aligned} \tag{6}$$

In Eq. (6), $f_1(Z^1_x, Z^2_y)$ denotes the expected shared surplus of brand 1 when brand 1 chooses strategy x and brand 2 selects strategy y. $f_2(Z^1_x, Z^2_y)$ is that of brand 2 under the same strategies. In the equation, U^b_{ij}/C^b_i is the segment j's utility for product i of brand b per unit cost. Here, the utility U^b_{ij} is computed based on Eq. (1) with weights from Eq. (4). The next term computes the market share that product i of brand b has in segment j, where μ is a scaling parameter, and I_{com} is the number of competing products under (Z^1_x, Z^2_y). Q_j is the market size of segment j.

Finally, the method discovers NE in the payoff matrix. NE, proposed by John Nash in the 1950s, provides robust solutions for competing players by mathematically modeling interactive strategic decision situations [39,40]. The concept can be defined as a profile of strategies such that each player's strategy is an optimal response to the other player's strategy [41]. This study considers pure Nash equilibrium (PNE) only because the study focuses on the product portfolio strategy. When a company determines its product portfolio, it cannot select two different portfolios due to limited budgets for product development. PNE allows only one strategy for a player, whereas NE allows multiple strategies with a probability distribution. The intuition behind PNE is to find the strategy that is optimal for both brands under competition. In the example of Table 2, the best strategy for brand 1 alters by the strategy of brand 2. If brand 2 chooses Z^2_1, brand 1 will choose Z^1_2 because it gives a higher payoff (0.3 instead of 0.2). When brand 2 selects Z^2_2, brand 1 will go with Z^1_1 for the higher payoff (0.3 instead of 0.1). In the same manner, brand 2 will choose Z^2_1 when brand 1 plays with Z^1_1 or Z^1_2. The case where both brands select their best strategy is (Z^1_2, Z^2_1), so it is PNE. Mathematically, Eq. (7) shows the conditions for the PNE.

$$\begin{aligned} x^T A y &\geq \tilde{x}^T A y \quad \forall\ \tilde{x} \in R_1 \\ x^T B y &\geq x^T B \tilde{y} \quad \forall\ \tilde{y} \in R_2 \end{aligned} \tag{7}$$

A is the payoff matrix for brand 1, and B is that for brand 2. In this equation, x and y are unit vectors indicating strategy choice. For example, when a brand has four feasible strategies, [0,1,0,0] indicates that the brand chooses the second strategy. The first equation means that the selected x gives the best payoff for brand 1 among all the possible strategies in R_1 when brand 2 chooses strategy y. Likewise, the second equation means the selected y provides the best payoff for brand 2 given that brand 1 chooses strategy x.

Customer	T_1	T_2	T_3	T_4	T_5	T_6	# Topics (T) mentioned
1	1	0	0	1	0	1	3
2	1	1	0	0	1	1	4
3	0	0	0	1	0	1	2

Fig. 3 Network construction process [29]

If the size of the payoff matrix is small, PNE can be found manually. However, finding PNE becomes complex when the matrix size gets large. This study implements a simple algorithm to identify PNE efficiently. The pseudo-code for finding PNE is shown in Algorithm 1. Z_a^b indicates that brand b chooses strategy a, and the algorithm prints out the optimal strategy (x, y), which leads to PNE.

Algorithm 1 Pure Nash equilibrium

```
f_1: Payoff for brand 1
f_2: Payoff for brand 2
Z_x^1: Brand 1 chooses strategy x
Z_y^2: Brand 2 chooses strategy y
for y = 1, 2, ..., N do
    Find x that maximizes f_1(Z_x^1, Z_y^2)
    Find k that maximizes f_2(Z_x^1, Z_k^2)
    if k = y then
        Print (x, y)
    end if
end for
```

The PNE will be different when the brand factor is considered and when it is not. Algorithm 1 shows that PNE depends on the payoff $f_b(Z_x^1, Z_y^2)$, and Eq. (6) shows that the payoff depends on customer utility U_{ij}^b. The reflection of the brand factor gives different customer utility for each brand, thus altering the payoff matrix and PNE. This change will be demonstrated in the following section.

3.4 Brand Effect Analysis. This study suggests a comparison model to analyze the brand effect in product portfolio strategies. The comparison model consists of two payoff matrices and PNE of them. One matrix represents the result with brand consideration, and another is without brand consideration. In specific, the method generates matrices in the format shown in Table 2, using Eq. (6). When the brand is considered, customer utility for brand 1 (U_{ij}^1) is calculated based on the TF of brand 1 reviews (TF_{jk}^1) according to Eq. (5), and the utility for brand 2 (U_{ij}^2) is computed based on the TF of brand 2 reviews (TF_{jk}^2). Without brand consideration, the two brands calculate the customer utility using the same TF, i.e., the TF of the total reviews. As a result, the method produces different payoff matrices, and they may have different PNE since the payoff values have changed. This study compares two portfolios corresponding to PNE and investigates how companies choose non-optimal strategies without brand consideration. Section 5 will demonstrate the comparison model with case study results.

Table 2 A payoff matrix

	Z_1^2	Z_2^2
Z_1^1	(0.2, 0.4)	(0.3, 0.1)
Z_2^1	(0.3, 0.5)	(0.1, 0.4)

4 Case Study

The proposed methodology was tested on real-world data. A smartphone was chosen as the target product because the market fits the game setting presented in Sec. 3. Specifically, the US smartphone market is dominated by two major brands—Apple and Samsung [42]. Also, the bill of materials (BoMs) can be easily obtained online [43].

4.1 Customer Analysis. In this section, the customer analysis in Fig. 1 was tested on smartphone review data, and the result is presented.

4.1.1 Data Collection and Preprocessing. The review data were collected from Amazon. Among the top 100 items in the cellphone category, 85 products were selected (non-smartphones and duplicated items were excluded). The data contain 44,691 reviews from Jul. 10, 2017 to Mar. 24, 2022. Only the reviews marked as "Verified Purchase" and written in the USA were considered. The collected reviews went through preprocessing. Symbols, numbers, and punctuation marks except a period were removed. Upper cases were converted to lower cases, and then all the words in the review data were lemmatized.

4.1.2 Feature Extraction. Product features of customer interests were extracted from the collected review data based on the methodology in Ref. [28]. In this case study, seven feature categories were detected: screen, application processor (AP), memory, camera, battery, unlock, and price. Table 3 shows the corresponding cue phrases for each feature category.

4.1.3 Customer Segmentation. Next, customers' interests and sentiments for product features were analyzed based on these cue phrases. Reviews not mentioning any product features were removed. The number of filtered customers was 13,961. Then, people with similar interests were connected by the networking rule presented in Ref. [29].

Figure 4 shows an example of the customer network. Modularity clustering was applied to this network and divided the customer base into different segments. In Fig. 4, segments are distinguished by colors. Figure 5 shows the characteristics of each segment. The x-axis indicates the feature and sentiment. For example, S+ means the positive sentiment for the screen feature, and B − means the negative sentiment for the battery feature. The y-axis represents the percentage of customers expressing each sentiment for features. People in different segments have different properties, i.e., interests and sentiments for product features. Specifically, customers in segment 1 have complaints about overall features. On the other hand, customers

Table 3 Cue phrases for product features

Screen	Screen display, screen size, inch display, screen resolution, screen brightness, screen sensitivity, screen ratio, lcd screen, oled screen, screen clarity, huge screen, large screen, big screen, screen edge, curved screen, etc.
AP	Fast processor, slow processor, snapdragon processor, exynos processor, process speed, processing speed
Memory	Gb memory, storage capacity, internal memory, more memory, extra memory, expandable memory, gb ram, more storage, enough space, great storage, extra storage, internal storage, storage space, gb storage, etc.
Camera	Front camera, selfie camera, rear camera, main camera, mp camera, camera lens, camera quality, camera app, camera function, camera software, camera upgrade, well camera, camera shutter, camera sound, etc.
Battery	Battery capacity, mah battery, battery charge, battery life, battery percentage, battery saver, battery health, battery power, battery replacement, replaceable battery, removable battery, battery drain, low battery, etc.
Unlock	Fingerprint reader, fingerprint sensor, fingerprint scanner, fingerprint reading, fingerprint recognition, finger scanner, finger reader, finger sensor, iris scanner, same finger, face recognition, facial recognition
Price	Price range, price difference, price tag, decent price, affordable price, awesome price, perfect price, cheap price, half price, retail price, amazing price, price drop, sale price, discount price, fair price, etc.

Note: AP: Application processor.

in segment 2 are satisfied with most features. In segment 3, people are interested in the battery feature only. In segment 4, most people care about the price. Customers in segment 5 have a high interest in the screen feature. In most segments, customers have positive opinions about product features. This is because online reviews bias toward positive sentiments along with high ratings.

4.1.4 Brand Analysis. Since segments have different properties, it can be inferred that they have different partial utilities, i.e., the utility for each feature category. Table 4 shows the partial utility calculated by Eq. (4). Since the US smartphone market is dominated by Apple and Samsung, this study analyzed the utility of these two brands. The result shows that partial utilities vary by brand. "Total" in the brand column represents the baseline for brand analysis. The utility values for "Total" were calculated based on the entire reviews for 85 products from eight brands (Apple, Samsung, Google, Motorola, OnePlus, ZTE, TCL, and BLU). There was no consideration for brands in this result. In Sec. 5, this paper demonstrates how the optimal product portfolio changes when the brand is considered in user data analysis.

4.2 Product Analysis. In the game setting of Table 1, each brand has four product candidates. Therefore, four sets of spec configurations need to be determined. As a preliminary work for this, available spec values for each product feature were defined, as shown in Table 5. For the simplicity of the simulation, this study considers one sub-feature for each feature category. For example, a feature category "screen" has multiple sub-features, such as screen size, screen resolution, screen type, etc. Among them, this study considers screen size only. Regarding other feature categories, the target sub-features are AP speed, memory ROM, number of rear cameras, battery capacity, unlock type, and price level. The costs of spec options were estimated based on the BoM of Samsung smartphones [43,44]. Table 6 shows an example of the BoM. It provides detailed specifications of components and corresponding costs.

This study configured the specifications of product candidates (P_i^b) in Table 1 by randomly selecting spec options from Table 5. The spec configuration should be realistic, so we confirmed that the suggested candidates are reproducible by comparing them to the smartphones previously released in the market. The three scenarios mentioned in Sec. 3.2 were tested based on the data obtained in the previous sections. Specifically, Table 7 shows the game setting for case 1. The utility by segment (U_{ij}^b) was filled out based on the partial utility from Sec. 4.1 and the specs of product candidates. Since each feature category has a different scale, the spec values were normalized by the min_max_scaling. In other words, the spec values range from 0 to 1. For example, the screen size of 6.9 in. becomes 1, and the size of 5.6 in. becomes 0. These scaled values were the input data for x_{ik}^b in Eq. (1). Because

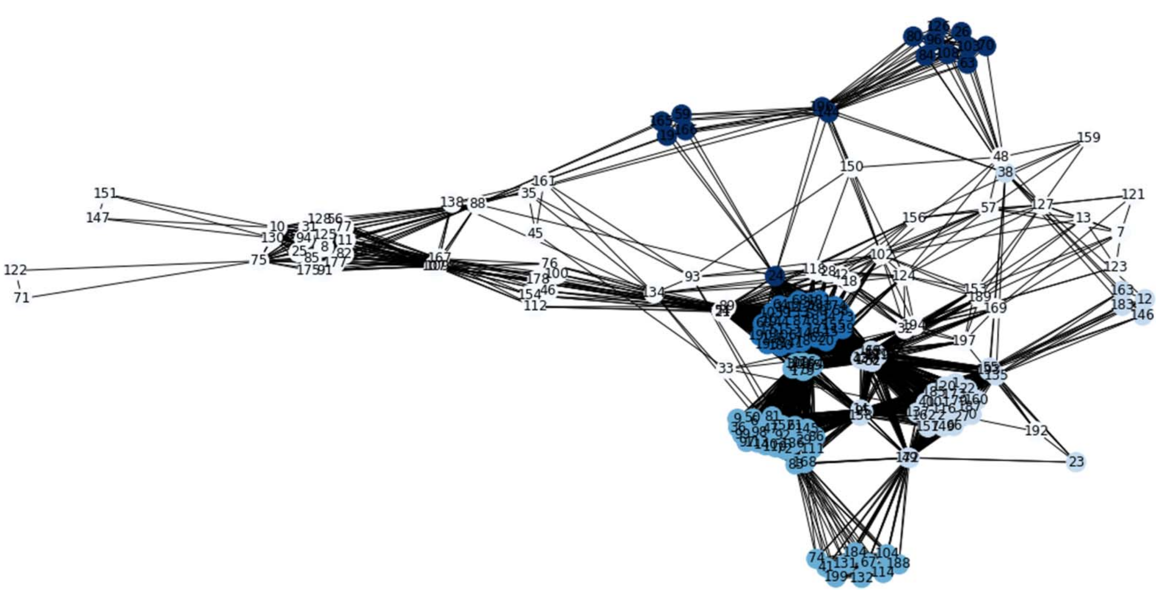

Fig. 4 A customer network with modularity clustering (*N* = 200)

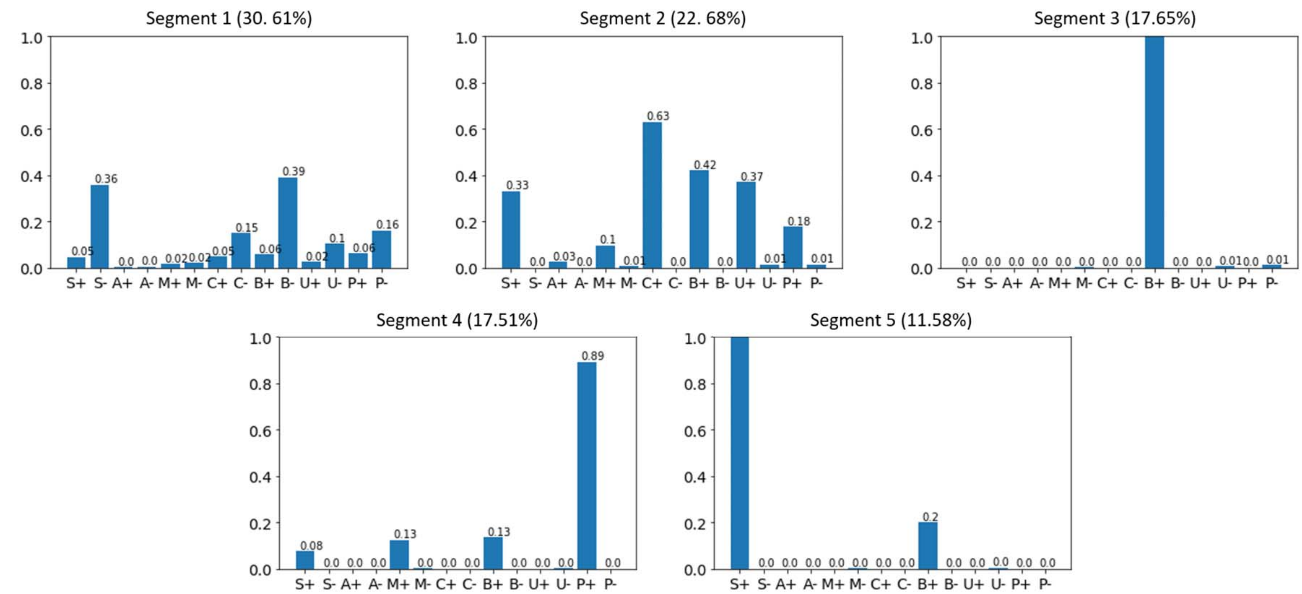

Fig. 5 Properties of customer segments

Table 4 Partial utility by segment

Brand	Segment	Screen	AP	Memory	Camera	Battery	Unlock	Price
Apple	1	0.205	0.004	0.043	0.137	0.258	0.096	0.258
	2	0.179	0.003	0.038	0.276	0.229	0.199	0.077
	3	0.001	0.001	0.001	0.001	0.997	0.001	0.001
	4	0.058	0.001	0.093	0.001	0.122	0.001	0.725
	5	0.825	0.001	0.001	0.001	0.172	0.001	0.001
Samsung	1	0.195	0.017	0.078	0.212	0.174	0.084	0.238
	2	0.153	0.019	0.055	0.331	0.175	0.173	0.094
	3	0.002	0.002	0.002	0.002	0.988	0.002	0.002
	4	0.070	0.001	0.105	0.001	0.074	0.001	0.747
	5	0.840	0.002	0.002	0.002	0.147	0.002	0.002
Total	1	0.181	0.011	0.060	0.189	0.221	0.096	0.242
	2	0.161	0.013	0.048	0.307	0.204	0.180	0.087
	3	0.000	0.000	0.000	0.000	0.998	0.000	0.000
	4	0.061	0.000	0.102	0.000	0.110	0.000	0.725
	5	0.830	0.001	0.001	0.001	0.167	0.001	0.001

Note: The utility values are summed up to 1 for each row.

Table 5 Spec options for product features

	Spec values				Cost ($)			
Feature	Option 1	Option 2	Option 3	Option 4	Option 1	Option 2	Option 3	Option 4
Screen size	5.6	6.0	6.4	6.9	61.5	71.5	81.5	91.5
AP speed	2.8	2.9	3.0	3.1	47.0	52.0	57.0	62.0
Memory ROM	64	128	256	512	56.5	61.5	66.5	71.5
Camera count	1	2	3	4	40.0	50.0	60.0	70.0
Battery capacity	3600	4000	4500	5000	4.5	7.2	9.9	12.6
Unlock type	0	1	2	3	10.0	13.0	16.0	19.0
Price	0	1	2	3	–	–	–	–

Table 6 Bill of materials for Galaxy Note20 Ultra 5G [43]

Category	Component	Specification	Cost
Display	Main display module + Touch + Driver integrated circuit (IC)	6.9 in., edge quad HD+, dynamic AMOLED, 496 ppi	$91.50
Processor	Application processor	Qualcomm snapdragon 865+	$57.00
Memory	NAND flash + Dynamic random access memory	128GB UFS + 12GB LPDDR5	$61.50
Camera	Module	12MP ultra wide angle, 108 MP wide angle, 12 MP telephoto	$60.30

Table 7 Game settings with data (case 1)

Players		Product (P_i^b)							Cost	Utility (U_{ij}^b)				
		S	A	M	C	B	U	P		S_1	S_2	S_3	S_4	S_5
Apple	P_1^1	6.4	3.1	512	4	3600	1	2	299.5	0.484	0.479	0.002	0.616	0.514
	P_2^1	5.6	2.8	256	2	4500	3	2	250.9	0.465	0.439	0.639	0.605	0.111
	P_3^1	6.0	2.8	64	1	5000	3	4	243.6	0.385	0.418	0.997	0.140	0.428
	P_4^1	6.9	2.8	64	1	4500	1	3	254.9	0.455	0.351	0.639	0.375	0.935
Samsung	P_1^2	6.4	3.1	512	4	3600	1	2	299.5	0.589	0.562	0.008	0.651	0.530
	P_2^2	5.6	2.8	256	2	4500	3	2	250.9	0.432	0.424	0.637	0.595	0.100
	P_3^2	6.0	2.8	64	1	5000	3	4	243.6	0.291	0.339	0.990	0.097	0.410
	P_4^2	6.9	2.8	64	1	4500	1	3	254.9	0.385	0.296	0.635	0.364	0.935

Note: S = screen size, A = AP speed, M = memory ROM, C = camera count, B = battery capacity, U = unlock type, P = price.

w/ Brand	Z_1^2	Z_2^2	Z_3^2	Z_4^2	Z_5^2	Z_6^2	Z_7^2	Z_8^2	Z_9^2	Z_{10}^2	Z_{11}^2	Z_{12}^2	Z_{13}^2	Z_{14}^2
Z_1^1	(0.069, 0.083)	(0.074, 0.095)	(0.078, 0.093)	(0.073, 0.098)	(0.047, 0.121)	(0.049, 0.125)	(0.047, 0.124)	(0.051, 0.125)	(0.049, 0.130)	(0.051, 0.126)	(0.037, 0.140)	(0.036, 0.141)	(0.037, 0.142)	(0.038, 0.143)
Z_2^1	(0.097, 0.085)	(0.095, 0.090)	(0.098, 0.087)	(0.098, 0.096)	(0.064, 0.119)	(0.065, 0.122)	(0.065, 0.123)	(0.065, 0.121)	(0.065, 0.127)	(0.066, 0.123)	(0.048, 0.137)	(0.048, 0.140)	(0.034, 0.103)	(0.049, 0.140)
Z_3^1	(0.104, 0.089)	(0.100, 0.092)	(0.097, 0.082)	(0.099, 0.094)	(0.070, 0.121)	(0.067, 0.120)	(0.069, 0.123)	(0.066, 0.119)	(0.067, 0.127)	(0.066, 0.120)	(0.051, 0.137)	(0.052, 0.140)	(0.051, 0.138)	(0.050, 0.139)
Z_4^1	(0.104, 0.085)	(0.105, 0.092)	(0.104, 0.085)	(0.100, 0.091)	(0.071, 0.120)	(0.069, 0.120)	(0.068, 0.121)	(0.071, 0.120)	(0.069, 0.125)	(0.068, 0.120)	(0.053, 0.137)	(0.052, 0.138)	(0.052, 0.138)	(0.052, 0.138)
Z_5^1	(0.114, 0.057)	(0.114, 0.062)	(0.117, 0.062)	(0.115, 0.066)	(0.085, 0.090)	(0.086, 0.093)	(0.086, 0.093)	(0.087, 0.093)	(0.086, 0.097)	(0.088, 0.095)	(0.069, 0.111)	(0.068, 0.113)	(0.069, 0.113)	(0.070, 0.114)
Z_6^1	(0.123, 0.058)	(0.121, 0.062)	(0.121, 0.059)	(0.121, 0.064)	(0.092, 0.091)	(0.091, 0.092)	(0.091, 0.093)	(0.091, 0.091)	(0.091, 0.096)	(0.091, 0.093)	(0.073, 0.111)	(0.073, 0.112)	(0.072, 0.112)	(0.072, 0.113)
Z_7^1	(0.119, 0.057)	(0.120, 0.063)	(0.121, 0.061)	(0.118, 0.063)	(0.090, 0.090)	(0.090, 0.092)	(0.088, 0.092)	(0.092, 0.093)	(0.090, 0.096)	(0.090, 0.093)	(0.073, 0.111)	(0.071, 0.112)	(0.072, 0.112)	(0.073, 0.113)
Z_8^1	(0.133, 0.059)	(0.132, 0.061)	(0.131, 0.056)	(0.132, 0.064)	(0.100, 0.090)	(0.099, 0.090)	(0.101, 0.093)	(0.099, 0.089)	(0.099, 0.095)	(0.099, 0.091)	(0.080, 0.109)	(0.080, 0.112)	(0.080, 0.111)	(0.080, 0.111)
Z_9^1	(0.135, 0.057)	(0.135, 0.061)	(0.136, 0.058)	(0.134, 0.062)	(0.102, 0.090)	(0.102, 0.090)	(0.101, 0.091)	(0.102, 0.090)	(0.101, 0.094)	(0.102, 0.091)	(0.082, 0.109)	(0.081, 0.111)	(0.081, 0.111)	(0.082, 0.111)
Z_{10}^1	(0.137, 0.059)	(0.137, 0.062)	(0.135, 0.055)	(0.134, 0.061)	(0.105, 0.091)	(0.103, 0.090)	(0.103, 0.092)	(0.103, 0.089)	(0.102, 0.094)	(0.101, 0.089)	(0.083, 0.109)	(0.083, 0.111)	(0.082, 0.110)	(0.082, 0.110)
Z_{11}^1	(0.140, 0.044)	(0.139, 0.046)	(0.140, 0.044)	(0.139, 0.049)	(0.112, 0.072)	(0.112, 0.073)	(0.112, 0.075)	(0.112, 0.073)	(0.111, 0.077)	(0.112, 0.075)	(0.093, 0.092)	(0.093, 0.094)	(0.093, 0.093)	(0.093, 0.094)
Z_{12}^1	(0.139, 0.043)	(0.140, 0.046)	(0.141, 0.046)	(0.139, 0.048)	(0.112, 0.072)	(0.112, 0.073)	(0.111, 0.074)	(0.113, 0.074)	(0.111, 0.076)	(0.112, 0.075)	(0.093, 0.092)	(0.093, 0.093)	(0.093, 0.093)	(0.094, 0.094)
Z_{13}^1	(0.143, 0.044)	(0.143, 0.047)	(0.143, 0.044)	(0.142, 0.047)	(0.115, 0.073)	(0.114, 0.073)	(0.114, 0.074)	(0.115, 0.073)	(0.114, 0.076)	(0.114, 0.073)	(0.096, 0.092)	(0.095, 0.093)	(0.095, 0.092)	(0.095, 0.093)
Z_{14}^1	(0.152, 0.044)	(0.152, 0.046)	(0.152, 0.042)	(0.151, 0.047)	(0.123, 0.072)	(0.122, 0.072)	(0.122, 0.074)	(0.122, 0.071)	(0.122, 0.075)	(0.121, 0.072)	(0.102, 0.091)	(0.102, 0.093)	(0.102, 0.092)	(0.102, 0.092)

w/o Brand	Z_1^2	Z_2^2	Z_3^2	Z_4^2	Z_5^2	Z_6^2	Z_7^2	Z_8^2	Z_9^2	Z_{10}^2	Z_{11}^2	Z_{12}^2	Z_{13}^2	Z_{14}^2
Z_1^1	(0.076, 0.076)	(0.079, 0.097)	(0.083, 0.100)	(0.079, 0.100)	(0.052, 0.118)	(0.053, 0.124)	(0.052, 0.120)	(0.055, 0.130)	(0.053, 0.132)	(0.055, 0.132)	(0.040, 0.140)	(0.039, 0.140)	(0.040, 0.142)	(0.041, 0.148)
Z_2^1	(0.097, 0.079)	(0.093, 0.093)	(0.095, 0.094)	(0.096, 0.099)	(0.063, 0.117)	(0.064, 0.121)	(0.065, 0.121)	(0.063, 0.127)	(0.063, 0.131)	(0.064, 0.129)	(0.047, 0.139)	(0.048, 0.139)	(0.034, 0.104)	(0.048, 0.146)
Z_3^1	(0.100, 0.083)	(0.094, 0.095)	(0.090, 0.090)	(0.092, 0.097)	(0.066, 0.119)	(0.063, 0.121)	(0.066, 0.121)	(0.062, 0.126)	(0.063, 0.131)	(0.061, 0.127)	(0.048, 0.138)	(0.049, 0.140)	(0.048, 0.140)	(0.046, 0.145)
Z_4^1	(0.100, 0.079)	(0.099, 0.096)	(0.097, 0.092)	(0.095, 0.095)	(0.068, 0.118)	(0.066, 0.121)	(0.065, 0.119)	(0.066, 0.127)	(0.065, 0.129)	(0.064, 0.127)	(0.050, 0.139)	(0.050, 0.138)	(0.049, 0.139)	(0.049, 0.145)
Z_5^1	(0.118, 0.052)	(0.117, 0.063)	(0.119, 0.066)	(0.118, 0.068)	(0.088, 0.088)	(0.089, 0.092)	(0.088, 0.091)	(0.089, 0.097)	(0.088, 0.099)	(0.089, 0.100)	(0.071, 0.112)	(0.070, 0.112)	(0.071, 0.114)	(0.071, 0.119)
Z_6^1	(0.124, 0.053)	(0.121, 0.064)	(0.121, 0.063)	(0.121, 0.066)	(0.092, 0.089)	(0.091, 0.091)	(0.092, 0.091)	(0.090, 0.096)	(0.090, 0.099)	(0.090, 0.097)	(0.073, 0.111)	(0.073, 0.112)	(0.073, 0.113)	(0.072, 0.117)
Z_7^1	(0.120, 0.052)	(0.121, 0.065)	(0.121, 0.066)	(0.119, 0.065)	(0.091, 0.088)	(0.091, 0.092)	(0.089, 0.089)	(0.091, 0.098)	(0.090, 0.098)	(0.090, 0.098)	(0.073, 0.112)	(0.072, 0.111)	(0.072, 0.113)	(0.073, 0.118)
Z_8^1	(0.130, 0.055)	(0.127, 0.063)	(0.126, 0.062)	(0.127, 0.066)	(0.097, 0.089)	(0.096, 0.090)	(0.098, 0.091)	(0.095, 0.095)	(0.095, 0.098)	(0.095, 0.097)	(0.077, 0.111)	(0.078, 0.112)	(0.077, 0.112)	(0.076, 0.117)
Z_9^1	(0.132, 0.053)	(0.131, 0.063)	(0.131, 0.063)	(0.129, 0.065)	(0.099, 0.088)	(0.099, 0.090)	(0.098, 0.090)	(0.098, 0.095)	(0.097, 0.097)	(0.097, 0.096)	(0.079, 0.111)	(0.079, 0.111)	(0.078, 0.112)	(0.078, 0.116)
Z_{10}^1	(0.132, 0.055)	(0.129, 0.064)	(0.127, 0.061)	(0.127, 0.064)	(0.100, 0.089)	(0.097, 0.090)	(0.098, 0.090)	(0.097, 0.095)	(0.096, 0.097)	(0.095, 0.095)	(0.079, 0.111)	(0.079, 0.111)	(0.077, 0.111)	(0.077, 0.116)
Z_{11}^1	(0.140, 0.040)	(0.139, 0.047)	(0.138, 0.048)	(0.139, 0.050)	(0.112, 0.071)	(0.111, 0.073)	(0.112, 0.073)	(0.111, 0.077)	(0.111, 0.079)	(0.111, 0.079)	(0.093, 0.093)	(0.093, 0.093)	(0.093, 0.094)	(0.092, 0.098)
Z_{12}^1	(0.140, 0.039)	(0.139, 0.048)	(0.140, 0.049)	(0.138, 0.050)	(0.112, 0.070)	(0.112, 0.073)	(0.111, 0.072)	(0.112, 0.078)	(0.111, 0.079)	(0.111, 0.079)	(0.093, 0.093)	(0.093, 0.093)	(0.093, 0.094)	(0.093, 0.098)
Z_{13}^1	(0.142, 0.040)	(0.141, 0.048)	(0.140, 0.048)	(0.139, 0.049)	(0.114, 0.071)	(0.113, 0.073)	(0.113, 0.072)	(0.112, 0.077)	(0.112, 0.078)	(0.111, 0.077)	(0.094, 0.093)	(0.094, 0.093)	(0.093, 0.093)	(0.093, 0.097)
Z_{14}^1	(0.148, 0.041)	(0.146, 0.048)	(0.145, 0.046)	(0.145, 0.049)	(0.119, 0.071)	(0.117, 0.072)	(0.118, 0.073)	(0.117, 0.076)	(0.116, 0.078)	(0.116, 0.077)	(0.098, 0.092)	(0.098, 0.093)	(0.097, 0.093)	(0.097, 0.097)

Fig. 6 Comparison of payoff matrix with/without brand consideration—case 1

w/ Brand	Z_1^2	Z_2^2	Z_3^2	Z_4^2	Z_5^2	Z_6^2	Z_7^2	Z_8^2	Z_9^2	Z_{10}^2	Z_{11}^2	Z_{12}^2	Z_{13}^2	Z_{14}^2
Z_1^1	(0.076, 0.086)	(0.084, 0.087)	(0.082, 0.093)	(0.087, 0.093)	(0.054, 0.120)	(0.053, 0.122)	(0.055, 0.128)	(0.056, 0.122)	(0.060, 0.118)	(0.058, 0.124)	(0.041, 0.138)	(0.042, 0.139)	(0.042, 0.142)	(0.044, 0.136)
Z_2^1	(0.093, 0.092)	(0.089, 0.080)	(0.093, 0.093)	(0.091, 0.085)	(0.061, 0.119)	(0.063, 0.124)	(0.062, 0.126)	(0.061, 0.119)	(0.060, 0.112)	(0.062, 0.121)	(0.047, 0.137)	(0.046, 0.137)	(0.036, 0.103)	(0.046, 0.133)
Z_3^1	(0.095, 0.090)	(0.098, 0.085)	(0.093, 0.089)	(0.096, 0.087)	(0.064, 0.120)	(0.063, 0.122)	(0.063, 0.126)	(0.064, 0.120)	(0.066, 0.115)	(0.063, 0.120)	(0.048, 0.137)	(0.048, 0.138)	(0.047, 0.140)	(0.049, 0.134)
Z_4^1	(0.105, 0.094)	(0.099, 0.081)	(0.100, 0.091)	(0.097, 0.082)	(0.070, 0.121)	(0.070, 0.123)	(0.068, 0.125)	(0.067, 0.119)	(0.066, 0.111)	(0.066, 0.118)	(0.053, 0.137)	(0.051, 0.136)	(0.051, 0.139)	(0.050, 0.132)
Z_5^1	(0.118, 0.060)	(0.120, 0.056)	(0.120, 0.062)	(0.121, 0.061)	(0.089, 0.090)	(0.089, 0.092)	(0.090, 0.095)	(0.090, 0.090)	(0.091, 0.087)	(0.091, 0.093)	(0.071, 0.110)	(0.072, 0.110)	(0.072, 0.113)	(0.073, 0.108)
Z_6^1	(0.117, 0.059)	(0.121, 0.058)	(0.119, 0.061)	(0.122, 0.062)	(0.090, 0.090)	(0.088, 0.091)	(0.089, 0.095)	(0.091, 0.091)	(0.093, 0.089)	(0.091, 0.092)	(0.072, 0.110)	(0.073, 0.111)	(0.072, 0.113)	(0.074, 0.109)
Z_7^1	(0.129, 0.061)	(0.129, 0.056)	(0.128, 0.061)	(0.128, 0.059)	(0.097, 0.090)	(0.096, 0.092)	(0.096, 0.094)	(0.096, 0.090)	(0.097, 0.086)	(0.096, 0.090)	(0.077, 0.110)	(0.077, 0.110)	(0.077, 0.112)	(0.077, 0.107)
Z_8^1	(0.127, 0.062)	(0.127, 0.055)	(0.127, 0.061)	(0.127, 0.058)	(0.095, 0.090)	(0.096, 0.093)	(0.095, 0.094)	(0.095, 0.089)	(0.096, 0.085)	(0.096, 0.090)	(0.076, 0.110)	(0.076, 0.109)	(0.077, 0.112)	(0.077, 0.107)
Z_9^1	(0.130, 0.064)	(0.126, 0.053)	(0.129, 0.062)	(0.126, 0.055)	(0.098, 0.090)	(0.099, 0.094)	(0.097, 0.094)	(0.096, 0.089)	(0.095, 0.083)	(0.096, 0.089)	(0.079, 0.110)	(0.077, 0.109)	(0.079, 0.112)	(0.077, 0.106)
Z_{10}^1	(0.132, 0.063)	(0.132, 0.056)	(0.130, 0.060)	(0.130, 0.056)	(0.100, 0.091)	(0.100, 0.092)	(0.099, 0.094)	(0.099, 0.089)	(0.099, 0.085)	(0.098, 0.089)	(0.080, 0.110)	(0.080, 0.109)	(0.079, 0.111)	(0.079, 0.106)
Z_{11}^1	(0.138, 0.046)	(0.139, 0.042)	(0.139, 0.046)	(0.140, 0.046)	(0.110, 0.072)	(0.110, 0.074)	(0.111, 0.076)	(0.111, 0.072)	(0.112, 0.070)	(0.112, 0.074)	(0.092, 0.091)	(0.093, 0.092)	(0.092, 0.094)	(0.093, 0.090)
Z_{12}^1	(0.143, 0.047)	(0.143, 0.041)	(0.143, 0.046)	(0.143, 0.044)	(0.115, 0.072)	(0.115, 0.075)	(0.114, 0.075)	(0.114, 0.071)	(0.114, 0.068)	(0.115, 0.073)	(0.096, 0.091)	(0.095, 0.091)	(0.096, 0.094)	(0.096, 0.089)
Z_{13}^1	(0.144, 0.046)	(0.145, 0.043)	(0.143, 0.045)	(0.144, 0.044)	(0.116, 0.073)	(0.115, 0.074)	(0.115, 0.075)	(0.115, 0.072)	(0.116, 0.070)	(0.115, 0.072)	(0.096, 0.092)	(0.096, 0.092)	(0.096, 0.093)	(0.096, 0.090)
Z_{14}^1	(0.146, 0.048)	(0.145, 0.041)	(0.146, 0.046)	(0.145, 0.042)	(0.117, 0.073)	(0.118, 0.075)	(0.117, 0.075)	(0.117, 0.071)	(0.117, 0.067)	(0.117, 0.071)	(0.098, 0.092)	(0.098, 0.091)	(0.098, 0.093)	(0.097, 0.088)

w/o Brand	Z_1^2	Z_2^2	Z_3^2	Z_4^2	Z_5^2	Z_6^2	Z_7^2	Z_8^2	Z_9^2	Z_{10}^2	Z_{11}^2	Z_{12}^2	Z_{13}^2	Z_{14}^2
Z_1^1	(0.080, 0.080)	(0.087, 0.089)	(0.085, 0.095)	(0.089, 0.100)	(0.056, 0.118)	(0.055, 0.120)	(0.057, 0.128)	(0.058, 0.124)	(0.061, 0.124)	(0.060, 0.129)	(0.043, 0.137)	(0.044, 0.141)	(0.043, 0.143)	(0.045, 0.142)
Z_2^1	(0.089, 0.087)	(0.084, 0.084)	(0.088, 0.096)	(0.085, 0.093)	(0.058, 0.118)	(0.060, 0.122)	(0.058, 0.127)	(0.057, 0.123)	(0.056, 0.119)	(0.058, 0.127)	(0.044, 0.138)	(0.043, 0.139)	(0.035, 0.103)	(0.043, 0.140)
Z_3^1	(0.095, 0.085)	(0.096, 0.088)	(0.092, 0.092)	(0.094, 0.094)	(0.064, 0.119)	(0.062, 0.120)	(0.062, 0.126)	(0.063, 0.124)	(0.064, 0.122)	(0.062, 0.126)	(0.047, 0.137)	(0.048, 0.140)	(0.047, 0.141)	(0.048, 0.140)
Z_4^1	(0.100, 0.089)	(0.093, 0.085)	(0.094, 0.094)	(0.090, 0.090)	(0.066, 0.119)	(0.067, 0.122)	(0.064, 0.126)	(0.063, 0.123)	(0.061, 0.118)	(0.062, 0.125)	(0.05, 0.138)	(0.048, 0.139)	(0.048, 0.141)	(0.047, 0.139)
Z_5^1	(0.118, 0.056)	(0.118, 0.058)	(0.119, 0.064)	(0.119, 0.066)	(0.088, 0.088)	(0.089, 0.091)	(0.089, 0.096)	(0.089, 0.093)	(0.090, 0.092)	(0.090, 0.097)	(0.071, 0.110)	(0.071, 0.112)	(0.072, 0.115)	(0.072, 0.113)
Z_6^1	(0.120, 0.055)	(0.122, 0.060)	(0.120, 0.062)	(0.122, 0.067)	(0.091, 0.089)	(0.090, 0.090)	(0.091, 0.096)	(0.091, 0.093)	(0.093, 0.094)	(0.091, 0.097)	(0.073, 0.110)	(0.073, 0.113)	(0.072, 0.114)	(0.074, 0.114)
Z_7^1	(0.128, 0.057)	(0.127, 0.058)	(0.126, 0.062)	(0.126, 0.064)	(0.096, 0.089)	(0.096, 0.091)	(0.095, 0.095)	(0.095, 0.092)	(0.095, 0.092)	(0.094, 0.095)	(0.076, 0.110)	(0.076, 0.111)	(0.076, 0.113)	(0.076, 0.112)
Z_8^1	(0.124, 0.058)	(0.123, 0.057)	(0.124, 0.063)	(0.123, 0.063)	(0.093, 0.089)	(0.093, 0.091)	(0.092, 0.095)	(0.092, 0.092)	(0.092, 0.091)	(0.092, 0.095)	(0.074, 0.110)	(0.074, 0.111)	(0.074, 0.114)	(0.074, 0.112)
Z_9^1	(0.124, 0.061)	(0.119, 0.056)	(0.122, 0.064)	(0.118, 0.061)	(0.092, 0.090)	(0.094, 0.093)	(0.092, 0.095)	(0.091, 0.092)	(0.089, 0.089)	(0.090, 0.095)	(0.075, 0.111)	(0.073, 0.111)	(0.074, 0.114)	(0.072, 0.111)
Z_{10}^1	(0.129, 0.060)	(0.127, 0.058)	(0.126, 0.062)	(0.125, 0.062)	(0.097, 0.090)	(0.097, 0.091)	(0.095, 0.094)	(0.095, 0.092)	(0.095, 0.090)	(0.094, 0.094)	(0.078, 0.110)	(0.077, 0.111)	(0.076, 0.113)	(0.076, 0.112)
Z_{11}^1	(0.137, 0.043)	(0.138, 0.044)	(0.137, 0.047)	(0.138, 0.050)	(0.110, 0.071)	(0.110, 0.073)	(0.110, 0.076)	(0.110, 0.074)	(0.111, 0.075)	(0.110, 0.078)	(0.091, 0.091)	(0.092, 0.093)	(0.092, 0.095)	(0.092, 0.094)
Z_{12}^1	(0.141, 0.044)	(0.139, 0.043)	(0.140, 0.048)	(0.139, 0.048)	(0.112, 0.071)	(0.113, 0.073)	(0.111, 0.076)	(0.111, 0.074)	(0.111, 0.073)	(0.111, 0.077)	(0.093, 0.092)	(0.093, 0.093)	(0.093, 0.095)	(0.093, 0.094)
Z_{13}^1	(0.143, 0.043)	(0.143, 0.045)	(0.141, 0.047)	(0.141, 0.048)	(0.115, 0.072)	(0.114, 0.072)	(0.113, 0.076)	(0.114, 0.074)	(0.114, 0.074)	(0.113, 0.076)	(0.095, 0.092)	(0.095, 0.093)	(0.094, 0.094)	(0.095, 0.094)
Z_{14}^1	(0.142, 0.045)	(0.140, 0.043)	(0.140, 0.048)	(0.139, 0.047)	(0.113, 0.072)	(0.114, 0.074)	(0.112, 0.076)	(0.112, 0.074)	(0.111, 0.072)	(0.112, 0.076)	(0.094, 0.092)	(0.094, 0.093)	(0.094, 0.095)	(0.093, 0.093)

Fig. 7 Comparison of payoff matrix with/without brand consideration—case 2

w/ Brand	Z_1^2	Z_2^2	Z_3^2	Z_4^2	Z_5^2	Z_6^2	Z_7^2	Z_8^2	Z_9^2	Z_{10}^2	Z_{11}^2	Z_{12}^2	Z_{13}^2	Z_{14}^2
Z_1^1	(0.099, 0.086)	(0.101, 0.098)	(0.103, 0.090)	(0.107, 0.091)	(0.067, 0.126)	(0.068, 0.121)	(0.070, 0.127)	(0.069, 0.128)	(0.071, 0.127)	(0.072, 0.122)	(0.051, 0.143)	(0.052, 0.145)	(0.053, 0.141)	(0.054, 0.142)
Z_2^1	(0.108, 0.089)	(0.103, 0.095)	(0.109, 0.090)	(0.108, 0.085)	(0.071, 0.126)	(0.074, 0.122)	(0.073, 0.126)	(0.071, 0.126)	(0.071, 0.123)	(0.074, 0.120)	(0.054, 0.142)	(0.054, 0.143)	(0.040, 0.103)	(0.054, 0.140)
Z_3^1	(0.097, 0.092)	(0.095, 0.101)	(0.095, 0.090)	(0.098, 0.089)	(0.064, 0.130)	(0.064, 0.124)	(0.064, 0.128)	(0.064, 0.130)	(0.065, 0.127)	(0.064, 0.122)	(0.048, 0.145)	(0.048, 0.147)	(0.048, 0.142)	(0.048, 0.143)
Z_4^1	(0.099, 0.096)	(0.092, 0.099)	(0.095, 0.092)	(0.091, 0.084)	(0.065, 0.131)	(0.067, 0.125)	(0.064, 0.128)	(0.063, 0.129)	(0.061, 0.124)	(0.063, 0.120)	(0.049, 0.145)	(0.048, 0.146)	(0.049, 0.142)	(0.047, 0.142)
Z_5^1	(0.139, 0.058)	(0.138, 0.064)	(0.141, 0.060)	(0.142, 0.060)	(0.104, 0.094)	(0.106, 0.090)	(0.107, 0.094)	(0.105, 0.095)	(0.106, 0.094)	(0.108, 0.091)	(0.085, 0.113)	(0.085, 0.115)	(0.086, 0.112)	(0.086, 0.113)
Z_6^1	(0.132, 0.059)	(0.131, 0.067)	(0.133, 0.060)	(0.135, 0.061)	(0.099, 0.096)	(0.099, 0.091)	(0.100, 0.095)	(0.099, 0.097)	(0.101, 0.096)	(0.102, 0.092)	(0.079, 0.115)	(0.080, 0.117)	(0.081, 0.113)	(0.081, 0.114)
Z_7^1	(0.137, 0.061)	(0.134, 0.066)	(0.136, 0.060)	(0.136, 0.058)	(0.102, 0.097)	(0.103, 0.092)	(0.103, 0.094)	(0.101, 0.096)	(0.102, 0.094)	(0.103, 0.090)	(0.082, 0.115)	(0.082, 0.116)	(0.083, 0.112)	(0.082, 0.113)
Z_8^1	(0.137, 0.061)	(0.134, 0.065)	(0.138, 0.060)	(0.138, 0.059)	(0.102, 0.096)	(0.104, 0.092)	(0.103, 0.094)	(0.102, 0.096)	(0.103, 0.094)	(0.104, 0.090)	(0.082, 0.114)	(0.082, 0.116)	(0.083, 0.112)	(0.083, 0.113)
Z_9^1	(0.137, 0.063)	(0.132, 0.064)	(0.136, 0.061)	(0.134, 0.056)	(0.102, 0.097)	(0.104, 0.093)	(0.102, 0.094)	(0.101, 0.095)	(0.100, 0.092)	(0.102, 0.090)	(0.083, 0.115)	(0.081, 0.115)	(0.083, 0.112)	(0.081, 0.112)
Z_{10}^1	(0.130, 0.064)	(0.127, 0.068)	(0.128, 0.060)	(0.128, 0.058)	(0.097, 0.099)	(0.098, 0.094)	(0.097, 0.096)	(0.096, 0.097)	(0.095, 0.095)	(0.096, 0.090)	(0.078, 0.117)	(0.077, 0.117)	(0.078, 0.113)	(0.077, 0.114)
Z_{11}^1	(0.153, 0.045)	(0.152, 0.049)	(0.154, 0.045)	(0.155, 0.045)	(0.122, 0.076)	(0.124, 0.073)	(0.124, 0.075)	(0.123, 0.076)	(0.124, 0.076)	(0.125, 0.073)	(0.103, 0.095)	(0.103, 0.096)	(0.104, 0.093)	(0.104, 0.094)
Z_{12}^1	(0.155, 0.046)	(0.153, 0.049)	(0.156, 0.045)	(0.155, 0.044)	(0.124, 0.077)	(0.126, 0.073)	(0.125, 0.075)	(0.124, 0.076)	(0.124, 0.075)	(0.126, 0.072)	(0.104, 0.095)	(0.104, 0.095)	(0.105, 0.093)	(0.104, 0.093)
Z_{13}^1	(0.150, 0.047)	(0.148, 0.051)	(0.150, 0.045)	(0.150, 0.045)	(0.120, 0.078)	(0.121, 0.074)	(0.120, 0.076)	(0.119, 0.077)	(0.120, 0.076)	(0.121, 0.072)	(0.100, 0.096)	(0.100, 0.097)	(0.101, 0.094)	(0.100, 0.095)
Z_{14}^1	(0.152, 0.048)	(0.148, 0.049)	(0.151, 0.045)	(0.150, 0.043)	(0.121, 0.078)	(0.122, 0.075)	(0.121, 0.075)	(0.120, 0.077)	(0.120, 0.075)	(0.121, 0.072)	(0.101, 0.096)	(0.101, 0.096)	(0.102, 0.094)	(0.100, 0.094)

w/o Brand	Z_1^2	Z_2^2	Z_3^2	Z_4^2	Z_5^2	Z_6^2	Z_7^2	Z_8^2	Z_9^2	Z_{10}^2	Z_{11}^2	Z_{12}^2	Z_{13}^2	Z_{14}^2
Z_1^1	(0.099, 0.078)	(0.099, 0.105)	(0.101, 0.090)	(0.105, 0.097)	(0.066, 0.126)	(0.068, 0.116)	(0.069, 0.125)	(0.068, 0.133)	(0.070, 0.135)	(0.070, 0.126)	(0.051, 0.143)	(0.052, 0.147)	(0.052, 0.140)	(0.052, 0.149)
Z_2^1	(0.109, 0.082)	(0.103, 0.103)	(0.109, 0.092)	(0.107, 0.092)	(0.071, 0.127)	(0.074, 0.118)	(0.073, 0.124)	(0.071, 0.132)	(0.070, 0.132)	(0.073, 0.125)	(0.054, 0.143)	(0.054, 0.146)	(0.041, 0.102)	(0.054, 0.148)
Z_3^1	(0.094, 0.084)	(0.092, 0.109)	(0.091, 0.091)	(0.093, 0.096)	(0.062, 0.130)	(0.062, 0.119)	(0.062, 0.127)	(0.061, 0.136)	(0.062, 0.137)	(0.062, 0.127)	(0.046, 0.145)	(0.046, 0.149)	(0.046, 0.141)	(0.046, 0.150)
Z_4^1	(0.102, 0.088)	(0.092, 0.107)	(0.096, 0.093)	(0.091, 0.091)	(0.066, 0.131)	(0.068, 0.121)	(0.065, 0.126)	(0.063, 0.135)	(0.061, 0.134)	(0.063, 0.126)	(0.050, 0.146)	(0.048, 0.148)	(0.049, 0.141)	(0.047, 0.149)
Z_5^1	(0.140, 0.053)	(0.137, 0.069)	(0.140, 0.061)	(0.141, 0.064)	(0.104, 0.094)	(0.106, 0.087)	(0.106, 0.093)	(0.104, 0.099)	(0.105, 0.101)	(0.107, 0.094)	(0.084, 0.113)	(0.084, 0.117)	(0.086, 0.111)	(0.085, 0.118)
Z_6^1	(0.130, 0.054)	(0.128, 0.073)	(0.129, 0.060)	(0.131, 0.066)	(0.097, 0.096)	(0.098, 0.088)	(0.098, 0.094)	(0.097, 0.101)	(0.098, 0.103)	(0.099, 0.095)	(0.078, 0.115)	(0.078, 0.119)	(0.079, 0.112)	(0.079, 0.120)
Z_7^1	(0.138, 0.056)	(0.134, 0.071)	(0.136, 0.061)	(0.136, 0.063)	(0.102, 0.097)	(0.104, 0.088)	(0.103, 0.093)	(0.101, 0.100)	(0.101, 0.101)	(0.102, 0.094)	(0.082, 0.115)	(0.082, 0.118)	(0.082, 0.112)	(0.081, 0.119)
Z_8^1	(0.137, 0.056)	(0.132, 0.071)	(0.136, 0.061)	(0.135, 0.063)	(0.101, 0.096)	(0.103, 0.089)	(0.102, 0.094)	(0.100, 0.100)	(0.100, 0.101)	(0.102, 0.094)	(0.081, 0.115)	(0.081, 0.118)	(0.082, 0.112)	(0.081, 0.119)
Z_9^1	(0.139, 0.058)	(0.132, 0.070)	(0.137, 0.062)	(0.134, 0.061)	(0.103, 0.097)	(0.105, 0.090)	(0.103, 0.093)	(0.101, 0.100)	(0.100, 0.100)	(0.102, 0.094)	(0.083, 0.116)	(0.082, 0.118)	(0.083, 0.112)	(0.081, 0.118)
Z_{10}^1	(0.130, 0.059)	(0.125, 0.073)	(0.127, 0.062)	(0.126, 0.063)	(0.096, 0.099)	(0.097, 0.091)	(0.096, 0.095)	(0.094, 0.102)	(0.094, 0.102)	(0.094, 0.094)	(0.077, 0.117)	(0.076, 0.119)	(0.077, 0.113)	(0.075, 0.120)
Z_{11}^1	(0.152, 0.041)	(0.150, 0.053)	(0.152, 0.045)	(0.153, 0.049)	(0.121, 0.076)	(0.123, 0.070)	(0.123, 0.074)	(0.121, 0.080)	(0.122, 0.082)	(0.123, 0.076)	(0.101, 0.095)	(0.102, 0.098)	(0.103, 0.093)	(0.102, 0.099)
Z_{12}^1	(0.157, 0.042)	(0.153, 0.053)	(0.156, 0.046)	(0.155, 0.047)	(0.124, 0.077)	(0.126, 0.071)	(0.125, 0.074)	(0.124, 0.080)	(0.123, 0.080)	(0.125, 0.075)	(0.104, 0.095)	(0.104, 0.097)	(0.105, 0.093)	(0.103, 0.099)
Z_{13}^1	(0.150, 0.042)	(0.147, 0.055)	(0.148, 0.045)	(0.148, 0.048)	(0.119, 0.078)	(0.120, 0.071)	(0.119, 0.075)	(0.118, 0.081)	(0.118, 0.082)	(0.119, 0.075)	(0.099, 0.096)	(0.099, 0.099)	(0.100, 0.093)	(0.099, 0.100)
Z_{14}^1	(0.152, 0.044)	(0.148, 0.054)	(0.150, 0.046)	(0.149, 0.047)	(0.120, 0.078)	(0.122, 0.072)	(0.121, 0.075)	(0.119, 0.081)	(0.118, 0.081)	(0.120, 0.075)	(0.101, 0.097)	(0.100, 0.098)	(0.101, 0.093)	(0.099, 0.099)

Fig. 8 Comparison of payoff matrix with/without brand consideration—case 3

higher prices decrease customer utility, the price data were converted to $(1-P)$ and plugged into the utility function. Table 7 shows the calculated customer utility by segment. The result demonstrates that customer utility varies by brand for the same product.

4.3 Product Portfolio Strategy. In this study, the feasible strategies were defined as "up to three products out of four candidates." Therefore, each brand has 14 strategies as follows. Z_a^b indicates the ath strategy for brand b.

$Z_1^1 = \{P_1^1\}$	$Z_1^2 = \{P_1^2\}$
$Z_2^1 = \{P_2^1\}$	$Z_2^2 = \{P_2^2\}$
$Z_3^1 = \{P_3^1\}$	$Z_3^2 = \{P_3^2\}$
$Z_4^1 = \{P_4^1\}$	$Z_4^2 = \{P_4^2\}$
$Z_5^1 = \{P_1^1, P_2^1\}$	$Z_5^2 = \{P_1^2, P_2^2\}$
$Z_6^1 = \{P_1^1, P_3^1\}$	$Z_6^2 = \{P_1^2, P_3^2\}$
$Z_7^1 = \{P_1^1, P_4^1\}$	$Z_7^2 = \{P_1^2, P_4^2\}$
$Z_8^1 = \{P_2^1, P_3^1\}$	$Z_8^2 = \{P_2^2, P_3^2\}$
$Z_9^1 = \{P_2^1, P_4^1\}$	$Z_9^2 = \{P_2^2, P_4^2\}$
$Z_{10}^1 = \{P_3^1, P_4^1\}$	$Z_{10}^2 = \{P_3^2, P_4^2\}$
$Z_{11}^1 = \{P_1^1, P_2^1, P_3^1\}$	$Z_{11}^2 = \{P_1^2, P_2^2, P_3^2\}$
$Z_{12}^1 = \{P_1^1, P_2^1, P_4^1\}$	$Z_{12}^2 = \{P_1^2, P_2^2, P_4^2\}$
$Z_{13}^1 = \{P_1^1, P_3^1, P_4^1\}$	$Z_{13}^2 = \{P_1^2, P_3^2, P_4^2\}$
$Z_{14}^1 = \{P_2^1, P_3^1, P_4^1\}$	$Z_{14}^2 = \{P_2^2, P_3^2, P_4^2\}$

The payoff for each feasible strategy was calculated by Eq. (6), and the payoff matrix was constructed. Algorithm 1 for finding PNE was implemented in PYTHON and applied to this payoff matrix.

5 Results and Discussion

The goal of this study is to demonstrate the influence of the brand factor in data-driven design using online user data. While previous studies analyzed customer opinions without considering divergent customer perceptions of different brands, this study distinguishes customer opinions by brands. Specifically, it shows how the optimal product portfolio changes when the brand factor is considered in user data analysis. The suggested framework is to draw feature importance by brands, thus assigning different customer utility to products in different brands, and then compare two brands' product portfolios with and without brand consideration. The baseline model is the game setting with the partial utility of "Total" brands in Table 4. In this setting, two brands (Apple and Samsung) have the same importance for each product feature. The comparative model is a new game setting with partial utility reflecting brand effects. Specifically, the partial utility of "Apple" and "Samsung" in Table 4 is used. In this setting, the two brands have different weights for product features. The payoff matrices for the baseline and comparative models are constructed, and then the PNE of the two models are compared. As mentioned in Sec. 4.2, this study tests three cases. The product candidates for case 1 are shown in Table 7, and those of cases 2 and 3 are presented in Appendix.

5.1 Case 1. In the first case, two brands have identical product candidates ($P_i^1 = P_i^2$). Figure 6 shows two payoff matrices for this case. The upper one is the result when the brand is considered, i.e., the partial utilities of Apple and Samsung from Table 4 were reflected. The first column lists the feasible strategies of Apple, and the top row lists those of Samsung. The value in each cell is the payoff for two brands according to selected strategies. For example, the highlighted value (0.102, 0.093) means that Apple and Samsung get 0.102 and 0.093, respectively, when Apple chooses strategy 14 (Z_{14}^1) and Samsung selected strategy 12 (Z_{12}^2). The other matrix is in the same format and shows the result when the brand is disregarded, i.e., the partial utilities of "Total" from Table 4 were reflected. In both matrices, the highlighted cell indicates PNE, the optimal strategy for two brands. It is observed that the reflection of the brand factor changes the optimal strategy. Specifically, when the brand factor is disregarded, PNE is (Z_{14}^1, Z_{14}^2). Therefore, the product portfolio for Apple is $\{P_2^1, P_3^1, P_4^1\}$ and the product strategy for Samsung is $\{P_2^2, P_3^2, P_4^2\}$. On the other hand, when the brand factor is considered, PNE is (Z_{14}^1, Z_{12}^2). Samsung's portfolio is changed to $\{P_1^2, P_2^2, P_4^2\}$. This study assumes that the payoff with the brand factor is the true one. Therefore, the lack of consideration for the brand effect leads to Samsung choosing a strategy that is not optimal. In specific, Samsung chooses Z_{14}^2 with an actual payoff of 0.092 (in the upper table) instead of Z_{12}^2 with a payoff of 0.093.

5.2 Case 2. In the second case, two brands again have the same product candidates ($P_i^1 = P_i^2$) but with spec configurations different from case 1. The payoff matrices with and without the brand factor are shown in Fig. 7, and PNE is highlighted in the matrix. Considering the brand factor, PNE is (Z_{14}^1, Z_{13}^2). Therefore, the product portfolio for Apple is $\{P_2^1, P_3^1, P_4^1\}$ and the product strategy for

Samsung is $\{P_1^2, P_3^2, P_4^2\}$. When the brand factor is not considered, PNE is (Z_{13}^1, Z_{13}^2). In this case, Apple's product portfolio is changed to $\{P_1^1, P_3^1, P_4^1\}$. Therefore, disregarding the brand effect leads to Apple selecting a non-optimal strategy. Apple chooses Z_{13}^1 with an actual payoff of 0.096 (in the upper table) rather than Z_{14}^1 with a payoff of 0.098.

5.3 Case 3. In case 3, two brands have different product candidates ($P_i^1 \neq P_i^2$ for some i). The payoff matrices are shown in Fig. 8, where the PNE cell is highlighted. With the brand factor, PNE is (Z_{12}^1, Z_{12}^2). Therefore, the product portfolio for Apple is $\{P_1^1, P_2^1, P_4^1\}$ and the product strategy for Samsung is $\{P_1^2, P_2^2, P_4^2\}$. Without the brand factor, PNE is (Z_{12}^1, Z_{14}^2), and Samsung's portfolio is changed to $\{P_2^2, P_3^2, P_4^2\}$. Disregarding brands, Samsung results in choosing Z_{14}^2 with a true payoff of 0.093 (in the upper table) instead of Z_{12}^2 with a payoff of 0.095.

6 Conclusion and Future Works

This study focused on the neglected brand effect in data-driven design based on online user-generated data. Online data have been a popular resource for customer analysis due to its strength in time and cost-efficiency compared to conventional data collection methods such as surveys and interviews. However, previous studies utilizing online data disregarded brand effects while it is a significant factor in the industry. In the field, companies research various brand indexes to identify their strengths and weaknesses and devise proper strategies for market competition. Therefore, the brand factor needs to be taken into account in relevant research.

This paper proposed a game theory-based approach to investigate the influence of the brand in product strategy based on user-generated online data. The approach consists of three stages: (i) customer analysis, (ii) product analysis, and (iii) product portfolio analysis. In the first stage, the customer base was divided into segments based on the online review data. Then, the method analyzed each segment's partial utility for product features. In the second stage, the method defined spec options for each feature category and determined spec configurations for product candidates. Finally, the game setting in Table 1 that represents market competition was filled in based on the results from (i) and (ii). The feasible strategies and corresponding payoffs were established, and PNE was discovered. This study compared the resultant PNE with and without brand consideration. As discussed in Sec. 5, disregarding brand effects resulted in a company choosing a non-optimal strategy for its product portfolio. In all three cases presented in this paper, the brand factor altered PNE, the optimal strategy for companies. These results demonstrate the importance of the brand factor in data-driven design using online data. The proposed methodology can be applied to other product domains, such as laptops and headphones. The review data for these products are available online, and the market is dominated by a few brands. The result helps companies devise optimal strategies for their product line-up by reflecting market competition in the real world.

The proposed methodology has some limitations to be addressed. First, the feature importance is calculated based on the TF. As mentioned in Sec. 3.1, developing a new method for feature importance would be outside the scope of this paper. However, the TF-based approach is a bit simple and may not provide high accuracy. In future research, more advanced methods for extracting feature importance, such as regression [9] and neural networks [45], will be considered. Second, the quality of customer data may affect the result of the proposed method. While NE is a robust method based on game theory, the payoff matrix for NE is constructed based on customer data. Changes in data affect feature importance, thus altering the payoff matrix and NE. For example, this study might have removed some reviews that mention target product features with synonyms of feature keywords. If we detect feature reviews manually, the result will include more reviews and may give different feature importance. In future works, context-based review detection, e.g., sentence bidirectional encoder representations from transformers (BERT) [46], will be studied for more accurate customer data processing. Finally, the suggested method considers the competition between two brands only. Finding NE becomes more complex when more than two players exist in the game setting. In addition, when the companies have a higher number of feasible strategies, the construction of the payoff matrix gets more complicated. In future works, the above issues will be studied, and more diverse cases will be tested. Regarding managerial application, the partial utility obtained in Sec. 3.1 can be applied to design applications other than the product portfolio. For example, companies can adopt this result for designing a new product that gives maximum customer utility. The previous studies utilizing online data provided various design applications, and the feature weight by brand obtained in this study can be applied to those applications. The new application can further demonstrate the importance of the brand factor.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the finding of this article are obtainable from the corresponding author upon reasonable request.

Nomenclature

b = brand
i = product
j = customer segment
k = product feature
I_{com} = number of competing products under (Z_x^1, Z_y^2)
Q_j = market size of segment j
R_b = a set of feasible strategies for brand b
w_{jk}^b = weight that segment j has for feature k in brand b
x_{ik}^b = spec value for feature k of product i of brand b
C_i^b = cost of product i of brand b
U_{ij}^b = utility that segment j has for product i of brand b
Z_x^1 = brand 1 chooses strategy x
Z_y^2 = brand 2 chooses strategy y
$f_b(Z_x^1, Z_y^2)$ = payoff for brand b with strategy (x, y)
TF_{jk}^b = number of positive reviews about feature k among segment j in brand b
μ = scaling parameter

Appendix: Game Setting for Cases 2 and 3

Tables 8 and 9 show product candidates for each brand in cases 2 and 3, respectively.

Table 8 Game settings with data (case 2)

Brand		S	A	M	C	B	U	P
Apple	P_1^1	6.4	3.1	256	3	3600	1	4
	P_2^1	6.9	2.8	128	1	4500	2	1
	P_3^1	5.6	2.9	256	2	4500	3	3
	P_4^1	6.0	2.8	64	1	5000	3	1
Samsung	P_1^2	6.4	3.1	256	3	3600	1	4
	P_2^2	6.9	2.8	128	1	4500	2	1
	P_3^2	5.6	2.9	256	2	4500	3	3
	P_4^2	6.0	2.8	64	1	5000	3	1

Table 9 Game settings with data (case 3)

Brand		S	A	M	C	B	U	P
Apple	P_1^1	6.4	2.9	256	3	4000	1	4
	P_2^1	5.6	3.0	128	1	4500	3	2
	P_3^1	6.0	2.8	256	2	4500	3	3
	P_4^1	6.9	3.1	64	1	5000	3	1
Samsung	P_1^2	6.4	2.9	256	3	3600	1	4
	P_2^2	5.6	3.0	128	1	4500	3	2
	P_3^2	6.0	2.8	256	2	4500	3	3
	P_4^2	6.9	3.1	64	1	5000	3	1

References

[1] Oh, T. T., Keller, K. L., Neslin, S. A., Reibstein, D. J., and Lehmann, D. R., 2020, "The Past, Present, and Future of Brand Research," Mark. Lett., **31**(2), pp. 151–162.
[2] Aaker, D. A., 2012, "Win the Brand Relevance Battle and Then Build Competitor Barriers," Calif. Manage. Rev., **54**(2), pp. 43–57.
[3] Na, W. B., Marshall, R., and Keller, K. L., 1999, "Measuring Brand Power: Validating a Model for Optimizing Brand Equity," J. Prod. Brand. Manage., **8**(3), pp. 170–184.
[4] Forbes, 2020, "The World's Most Valuable Brands," https://www.forbes.com/powerful-brands/list/3/
[5] Businesswire, 2022, "Hp Sets Its Sights on Apple in the PC Industry, Customers Remain Dissatisfied With Household Appliances, ACSI Data Show," https://www.businesswire.com/news/home/20210921005181/en
[6] Help, S. T., 2022, "Top 10 Market Research Companies," https://www.softwaretestinghelp.com/market-research-companies/
[7] Rekettye, G., and Liu, J., 2001, "Segmenting the Hungarian Automobile Market Brand Using Perceptual and Value Mapping," J. Target. Meas. Anal. Mark., **9**(3), pp. 241–253.
[8] Kim, H. H. M., Liu, Y., Wang, C. C., and Wang, Y., 2017, "Data-Driven Design (D3)," ASME J. Mech. Des., **139**(11), p. 110301.
[9] Suryadi, D., and Kim, H., 2018, "A Systematic Methodology Based on Word Embedding for Identifying the Relation Between Online Customer Reviews and Sales Rank," ASME J. Mech. Des., **140**(12), p. 121403.
[10] Joung, J., and Kim, H. M., 2021, "Automated Keyword Filtering in Latent Dirichlet Allocation for Identifying Product Attributes From Online Reviews," ASME J. Mech. Des., **143**(8), p. 084501.
[11] Yang, B., Liu, Y., Liang, Y., and Tang, M., 2019, "Exploiting User Experience From Online Customer Reviews for Product Design," Int. J. Inf. Manage., **46**(Jun), pp. 173–186.
[12] Chaklader, R., and Parkinson, M. B., 2017, "Data-Driven Sizing Specification Utilizing Consumer Text Reviews," ASME J. Mech. Des., **139**(11), p. 111406.
[13] Park, S., Joung, J., and Kim, H., 2023, "Spec Guidance for Engineering Design Based on Data Mining and Neural Networks," Comput. Ind., **144**, p. 103790.
[14] Tuarob, S., and Tucker, C. S., 2015, "Automated Discovery of Lead Users and Latent Product Features by Mining Large Scale Social Media Networks," ASME J. Mech. Des., **137**(7), p. 071402.
[15] Goldberg, D. M., and Abrahams, A. S., 2022, "Sourcing Product Innovation Intelligence From Online Reviews," Decis. Support Syst., **157**, p. 113751.
[16] Bertoni, A., "Data-Driven Design in Concept Development: Systematic Review and Missed Opportunities," 16th International Design Conference, Cavtat, Croatia, Oct. 26–29, Vol. 1, Cambridge University Press, pp. 101–110.
[17] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J., 2013, "Distributed Representations of Words and Phrases and Their Compositionality," 27th Annual Conference on Neural Information Processing Systems, Lake Tahoe, NV, Dec. 5–10.
[18] Wang, M., and Chen, W., 2015, "A Data-Driven Network Analysis Approach to Predicting Customer Choice Sets for Choice Modeling in Engineering Design," ASME J. Mech. Des., **137**(7), p. 071410.
[19] Chen, W., Hoyle, C., and Wassenaar, H. J., 2012, *Decision-Based Design*, Springer, London.
[20] Blei, D. M., Ng, A. Y., and Jordan, M. I., 2003, "Latent Dirichlet Allocation," J. Mach. Learn. Res., **3**, pp. 993–1022.
[21] Roth, A. E., 1988, *The Shapley Value*, Cambridge University Press, Cambridge.
[22] Lundberg, S. M., and Lee, S., 2017, "A Unified Approach to Interpreting Model Predictions," Thirty-first Conference on Neural Information Processing Systems, Long Beach, CA, Dec. 4–9.
[23] Jin, J., Liu, Y., Ji, P., and Liu, H., 2016, "Understanding Big Consumer Opinion Data for Market-Driven Product Design," Int. J. Prod. Res., **54**(10), pp. 3019–3041.
[24] Tuarob, S., and Tucker, C. S., 2015, "Quantifying Product Favorability and Extracting Notable Product Features Using Large Scale Social Media Data," ASME J. Comput. Inf. Sci. Eng., **15**(3), p. 031003.
[25] Nuortimo, K., and Harkonen, J., 2019, "Establishing an Automated Brand Index Based on Opinion Mining: Analysis of Printed and Social Media," J. Mark. Anal., **7**(3), pp. 141–151.
[26] Alzate, M., Arce-Urriza, M., and Cebollada, J., 2022, "Mining the Text of Online Consumer Reviews to Analyze Brand Image and Brand Positioning," J. Retail. Consum. Serv., **67**, p. 102989.
[27] Sadeghi, A., and Zandieh, M., 2011, "A Game Theory-Based Model for Product Portfolio Management in a Competitive Market," Expert Syst. Appl., **38**(7), pp. 7919–7923.
[28] Park, S., and Kim, H. M., 2022, "Phrase Embedding and Clustering for Sub-Feature Extraction From Online Data," ASME J. Mech. Des., **144**(5), p. 054501.
[29] Park, S., and Kim, H. M., 2022, "Finding Social Networks Among Online Reviewers for Customer Segmentation," ASME J. Mech. Des., **144**(12), p. 121703.
[30] Farrell, R. S., and Simpson, T. W., 2008, "A Method to Improve Platform Leveraging in a Market Segmentation Grid for an Existing Product Line," ASME J. Mech. Des., **130**(3), p. 031403.
[31] Suryadi, D., and Kim, H. M., 2019, "A Data-Driven Methodology to Construct Customer Choice Sets Using Online Data and Customer Reviews," ASME J. Mech. Des., **141**(11), p. 111103.
[32] Hutto, C., and Gilbert, E., "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," The Eighth International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI, June 1–4.
[33] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E., 2008, "Fast Unfolding of Communities in Large Networks," J. Stat. Mech.: Theory Exp., **2008**(10), p. P10008.
[34] Dinh, T. N., Li, X., and Thai, M. T., 2015, "Network Clustering Via Maximizing Modularity: Approximation Algorithms and Theoretical Limits," 2015 IEEE International Conference on Data Mining, Atlantic City, NJ, Nov. 14–17.
[35] Kim, J., Park, S., and Kim, H. M., 2022, "Optimal Modular Remanufactured Product Configuration and Harvesting Planning for End-of-Life Products," J. Mech. Des., **144**(4), p. 042001.
[36] Lewis, K., and Mistree, F., "Collaborative, Sequential, and Isolated Decisions in Design," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Sacramento, CA, Sept. 14–17.
[37] Sha, Z., Kannan, K. N., and Panchal, J. H., 2015, "Behavioral Experimentation and Game Theory in Engineering Systems Design," ASME J. Mech. Des., **137**(5), p. 051405.
[38] Bayrak, A. E., and Sha, Z., 2021, "Integrating Sequence Learning and Game Theory to Predict Design Decisions Under Competition," ASME J. Mech. Des., **143**(5), p. 051401.
[39] Facchinei, F., and Pang, J.-S., 2010, "12 Nash Equilibria: The Variational Approach," *Convex Optimization in Signal Processing and Digital Communication*, Cambridge University Press, New York, pp. 443–493.
[40] Gabriel, S. A., Conejo, A. J., Fuller, J. D., Hobbs, B. F., and Ruiz, C., 2012, *Complementarity Modeling in Energy Markets*, Springer Science & Business Media, New York.
[41] Fudenberg, D., and Tirole, J., 1991, *Game Theory*, MIT Press, Cambridge, MA.
[42] Counterpoint, T., 2022, "US Smartphone Market Share: By Quarter," https://www.counterpointresearch.com/us-market-smartphone-share/
[43] GSMArena, 2020, "Samsung Galaxy Note20 Ultra 5G Bom," https://www.gsmarena.com/samsung˙galaxy˙note20˙ultra˙5g˙bom˙is˙550˙calculate˙analysts-news-45209.php
[44] Businesswire, 2018, "Galaxy S9+ Materials Cost $43 More Than Previous Versions, IHS Markit Teardown Shows," https://www.businesswire.com/news/home/20180321005364/en
[45] Joung, J., and Kim, H. M., 2021, "Approach for Importance—Performance Analysis of Product Attributes From Online Reviews," ASME J. Mech. Des., **143**(8), p. 081705.
[46] Reimers, N., and Gurevych, I., 2019, "Sentence-Bert: Sentence Embeddings Using Siamese Bert-Networks," preprint arXiv:1908.10084.