mathematics

MDPI

Article

A Multi-Stage Model for Perceived Quality Evaluation of Clothing Brands

Minhui Ren [1], Yu Fan [1,2], Jindong Chen [1,3,*] and Jian Zhang [1,2,3]

[1] School of Economics and Management, Beijing Information Science and Technology University, Beijing 100192, China; 2021020714@bistu.edu.cn (M.R.); fanyu@bistu.edu.cn (Y.F.)
[2] Beijing Key Laboratory of Big Data Decision Making for Green Development, Beijing Information Science and Technology University, Beijing 100192, China
[3] Beijing International Science and Technology Cooperation Base for Intelligent Decision and Big Data Application, Beijing 100192, China
* Correspondence: j.chen@bistu.edu.cn

Abstract: Perceived quality is crucial for the functioning of clothing brands. However, accurate evaluation of the perceived quality of clothing brands remains a common challenge. To achieve a multidimensional evaluation of the perceived quality of clothing brands, an index system is derived based on perceived quality theory. Then, by combining a fine-grained sentiment analysis approach with stochastic dominance criteria, a multi-stage model ECRM is proposed for the perceived quality evaluation of clothing brands based on online user reviews. ECRM comprises three stages: Extraction, Classification, and Ranking. To begin with, Contrastive Attention and dependency parsing are used to extract attribute–viewpoint phrases from online reviews. Subsequently, the pre-trained models are employed to classify the indexes and sentiment levels of these phrases. Furthermore, the perceived quality indexes are ranked using stochastic dominance criteria and the PROMETHEE-II method. Empirical analysis is conducted for the clothing brands of ALDB, AND, BNL, and QPL; the results show that, based on online user reviews, ECRM enables accurate evaluation of the perceived quality of clothing brands. Based on the evaluation results, it is found that Comfort, External, Protection, and Fineness are highly valued by consumers; moreover, the four brands focus on different indexes. Specific strategies for perceived quality improvements are proposed depending on the current status of the brands.

Keywords: perceived quality; fine-grained sentiment analysis; stochastic dominance criteria; pre-trained models

MSC: 05A05

check for updates

Citation: Ren, M.; Fan, Y.; Chen, J.; Zhang, J. A Multi-Stage Model for Perceived Quality Evaluation of Clothing Brands. *Mathematics* **2023**, *11*, 3928. https://doi.org/10.3390/math11183928

Academic Editor: José Antonio Sanz

Received: 1 August 2023
Revised: 6 September 2023
Accepted: 12 September 2023
Published: 15 September 2023

Copyright: © 2023 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

Clothing consumption is entering a rational stage of the pursuit of quality and perception [1]. Clothing brands with low perceived quality can hardly meet consumers' demands. How to evaluate the perceived quality has become a multi-attribute decision making problem (MADM) for clothing brands [2]. For the last few years, residents' online consumption habits have been consolidated in the modern business context, where online shopping for clothing prevails [3]. Clothing brands need a set of systems and methods to accurately identify the directions for perceived quality improvements.

The concept of perceived quality was first proposed in the 1980s, and many theoretical studies had been conducted, such as the theoretical system analysis of perceived quality, perceived value, and perceived effort [4], the role of perceived quality in brand research [5], and the mechanism of perceived quality on consumers [6]. Questionnaire surveys have been widely used in the study of perceived quality, but they suffer from difficulties in data collection and question design. User-generated content (UGC) has changed the way

consumers share their experiences [7], and the unstructured UGC data has provided more advantages for data collection.

One type of UGC data is online user reviews. With the development of the internet, e-commerce platforms have gathered numerous user online reviews, serving as significant support in the research of consumer viewpoints and inclinations [8,9]. Many studies on perceived quality have been conducted using online reviews and sentiment analysis [10–13]. However, sentence-level sentiment analysis methods focus on the sentiment polarities of the whole sentences, and the connections of attributes and sentiments are overlooked [13]. To dig deeper into consumer perspectives, fine-grained sentiment analysis methods are needed.

Fine-grained sentiment analysis tasks include Aspect Term Extraction, Opinion Term Extraction, Aspect Sentiment Classification, etc., [14–18]. In response to these tasks, some fine-grained sentiment analysis methods are proposed using word vectors and language models. Tulkens et al. put forward an unsupervised approach to extract and classify aspect terms using Contrastive Attention [19]. Tian et al. proposed a pre-trained model, Skep (Sentiment Knowledge Enhanced Pre-training), based on sentiment knowledge augmentation. Skep acquired a unified sentiment representation via a dedicated objective function during pre-training, and demonstrated exceptional proficiency for the Aspect-level Sentiment Classification task [20]. Fine-grained sentiment analysis methods have been widely used in the evaluation studies of satisfaction [21], user experience [13], and service quality [22].

After sentiment analysis, a barometer, such as mean or variance, is needed to quantify the index. However, using only two numbers to represent the evaluation results is unreliable. Therefore, it is necessary to apply a more stable quantitative approach. Based on the result of sentiment analysis, a probabilistic linguistic term set (PLTS) can be constructed [8]. PLTS describes the reviews according to certain indexes; then, the dominance relations can be obtained using stochastic dominance criteria [23]. PROMETHEE-II is a complete ranking method based on dominance relations. It introduces an advantage function to compute the priority of the indexes, and the overall evaluation is given by the net advantages [24]. The combination of stochastic dominance criteria and PROMETHEE-II has been widely used in MADM problems [23].

In summary, the study of perceived quality holds significant importance for clothing brands, but simple sentiment and quantitative analysis methods have proven inadequate. To address the limitations of previous analyses, this paper presents an index system designed for evaluating the perceived quality of clothing brands. Further, a multi-stage model ECRM is proposed, and empirical analysis is conducted. The contributions of this paper are as follows. First, the perceived quality theory is involved to construct the index system of perceived quality evaluation for clothing brands. Second, a multi-stage model ECRM is proposed with a stochastic multi-attribute mathematical model and a fine-grained sentiment analysis method. Third, considering the online user reviews, this paper provides a multidimensional quantitative basis for evaluating the perceived quality of clothing brands.

The remainder of this paper is organized as follows. In Section 2, we present and explain the index system of perceived quality evaluation for clothing brands. In Section 3, we introduce the proposed model ECRM in detail. In Section 4, we conduct empirical analysis and discuss the results. In Section 5, we summarize this article.

2. Index System of Perceived Quality Evaluation for Clothing Brands

The clothing industry plays a significant role in consumer goods, and the perceived quality of clothing brands is crucial for resident consumption. According to the manner of online transactions and the theoretical system of perceived quality, this paper sets two evaluation targets: products and services. The index system of perceived quality evaluation for clothing brands is shown in Table 1.

Table 1. Index system of perceived quality evaluation for clothing brands.

Evaluation Targets	Evaluation Indexes
Products	Comfort Durability Fineness Protection Certification Appearance External
Services	Timeliness of response Effectiveness of communication Security of information Rationality of compensation Professionalism of service

In the classical theory, perceived quality applies exclusively to products, and is determined by three factors: external attributes, internal attributes, and perceived monetary price [4]. This conceptual model provides an overall theoretical framework but is not industry-specific. As for the perceived quality of clothing products, Liu et al. derived seven factors using questionnaire surveys [25], including Comfort, Certification, External, Fineness, Mall, Appearance, and Protection. Mall is specific to offline shopping. As consumption patterns change, more and more consumers tend to buy clothes online. The share of online clothing consumption has exceeded that of offline consumption [1], so this paper has removed the Mall index.

Parasuraman et al. summarized customer perceived service quality as five aspects: Reliability, Responsiveness, Assurance, Empathy, and Tangibility [26]. Similar to the classical perceived quality theoretical system, the perceived service quality system can provide theoretical support, but the interpretation of the indexes cannot justify their application in industries. Based on the study of Parasuraman et al. [27], Zhao et al. developed an e-tailing service quality measurement scale (E-TAIL-SQ) using in-depth interviews and empirical testing methods [28]. This paper refers to the first-order indices of E-TAIL-SQ to determine the service indexes.

Therefore, according to previous studies on perceived quality, the index system of perceived quality evaluation for clothing brands has been devised, and the indexes are determined as follows.

- Comfort: The index measures the comfort of the clothing items, considering aspects such as comfort, symmetry, and precision of sizing.
- Durability: Eugene [29] considers durability as a dimension of perceived product quality. Durability in this paper means the deformation of clothing products in daily use, including resistance to shrinkage and wrinkling of clothing.
- Fineness: Eugene [29] and Dodds [30] argue that product perceived quality should include fineness and technical level, such as dimensions. Fineness in this paper includes threads, seams, fabric breakage, pilling, and other related factors.
- Protection: Reliability should be considered as a dimension that reflects factors related to consumer safety [30], such as the color fastness or the fabric type of clothing. These factors are measured as aspects of Protection.
- Certification: Certification is employed to measure the compliance of signage.
- Appearance: Appearance is employed to measure the visual characteristics of the products, such as color or design.
- External: External is utilized to measure aspects that are not directly connected to product performance or physical components, such as brand or pricing.
- Timeliness of response: This index is derived from SERVQUAL and E-TAIL-SQ, and it is utilized to evaluate the promptness of customer service responses and punctuality of delivery.

- Effectiveness of communication: Whether the customer service is efficient and amiable in communication.
- Security of information: Whether the protection of consumers' personal information is ensured.
- Rationality of compensation: This index quantifies the rationality of the services provided during the return and exchange procedures.
- Professionalism of service: Whether the logistician can complete the distribution successfully and the retailer can process the order accurately.

3. The Multi-Stage Model for Perceived Quality Evaluation

Based on fine-grained sentiment analysis, a multi-stage model ECRM for perceived quality evaluation is proposed, where the output of each stage serves as the input for the next stage. The flowchart for ECRM is shown in Figure 1. First, the attribute–viewpoint phrases are extracted by Contrastive Attention and Dependency Parsing. Then, based on the evaluation index system, the sentiment levels and indexes of reviews are classified using the phrases. Finally, according to PLTS formed by the classification results, the indexes are ranked using stochastic dominance criteria and PROMETHEE-II.

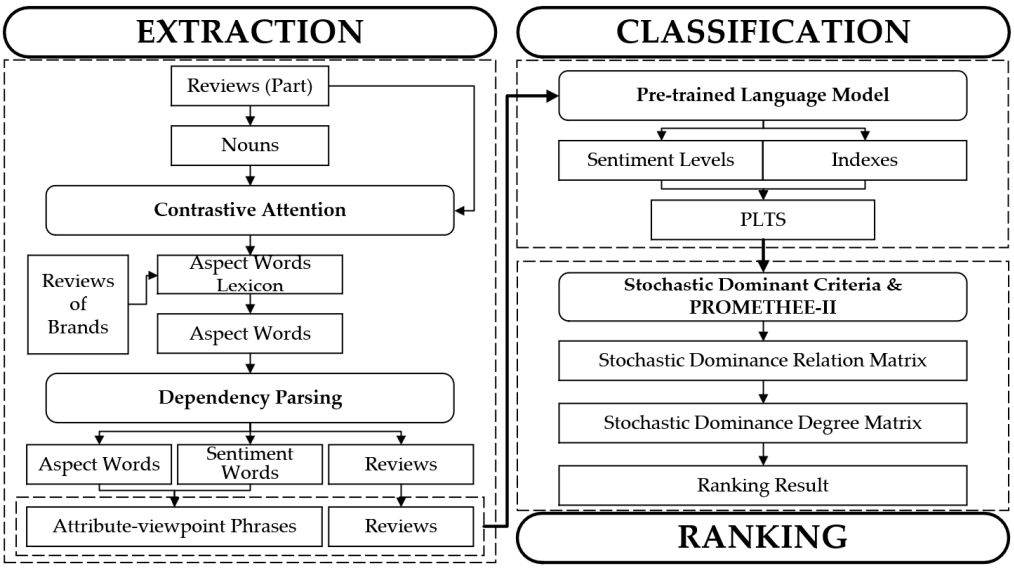


Figure 1. Flowchart for ECRM.

3.1. Extraction of Attribute–Viewpoint Phrases

In the pipeline style approaches of fine-grained sentiment analysis, the extraction of key words has always been the initial stage. In this paper, aspect words and sentiment words are extracted and combined into attribute–viewpoint phrases. The process can be described as follows.

Step 1: Obtain all the terms from the corpus by Chinese word segmentation.

Step 2: Since nouns typically provide information on attributes, select all nouns in the corpus as candidate aspect words [21].

Step 3: Select aspect words from the nouns using Contrastive Attention [21]. First, BERT is applied to obtain word embeddings [31]. Then, the Radial Basis Function (RBF) is used to compute the similarity between terms. RBF is defined as (1), where γ is a scaling factor.

$$\mathrm{RBF}(x, y, \gamma) = \exp\left(-\gamma \|x - y\|_2^2\right) \tag{1}$$

The kernel functions can be used to calculate the inner product of samples mapped to a higher dimensional space. Thus, the kernel functions can be utilized to measure the similarity between samples. RBF is one of the kernel functions. The active range of RBF can

be regulated by a single parameter. RBF can normalize the distance in the vector space to (0,1]. If the embeddings of x and y are highly similar, the RBF response will be close to 1; as the distance between x and y increases, the RBF response decreases rapidly until almost 0.

Let N be the word embeddings of all the candidate aspect words, and let M be the embeddings of the corpus. Compute the attention scores between nouns and the whole corpus as (2).

$$Att = \frac{\sum_{n \in N} \mathrm{RBF}(m, n, \gamma)}{\sum_{m \in M} \sum_{n \in N} \mathrm{RBF}(m, n, \gamma)} \tag{2}$$

Att represents the possibility of each noun being an aspect word. Equation (2) means the sum of the RBF distance of each noun to the corpus divided by the sum of the RBF distance of all nouns to the corpus, and high Att means high similarity between nouns and the corpus.

Finally, retain nouns with high Contrastive Attention scores using the elbow method.

Step 4: Extract attribute–viewpoint phrases from online reviews based on dependency parsing. Dependency parsing is a commonly used method for extracting the sentiment words related to the aspect words [32]. Since the aspect words have been determined, it is feasible to extract the viewpoints of online user reviews based on dependency patterns. First, the Part-Of-Speech (POS) of words in online reviews is annotated by Jieba https://github.com/fxsjy/jieba accessed on (accessed on 13 July 2023). Subsequently, based on prior research [26,27], five rules in dependency parsing are defined and displayed in Table 2.

Table 2. Extraction rules for attribute–viewpoint phrases.

Extraction Rules	Examples of Reviews	Extraction Results
aspect + adj	The shirt is nice.	shirt nice
adj + aspect	The color is as beautiful as snow.	color beautiful
aspect + adv + adj	The delivery is very fast.	delivery very fast
adv + adj + aspect	It's the most beautiful dress I've ever had.	most beautiful dress
verb + aspect	I like that material.	like material

There are six relations in dependency parsing: subject–verb relation, verb–object relation, fronting–object relation, attribute relation, adverbial relation, complement relation, and coordinate relation. The rules defined in Table 2 can address the format of the first five relations directly. As for aspect words of coordinate relation, the sentences can be broken down into separate aspect words and then analyzed by the first five rules [32,33].

It can be seen that content words are sometimes not linked in raw reviews; however, during Chinese word segmentation, insignificant words are eliminated, and content words are retained in their original order.

3.2. Classification of Reviews

The pre-training approach has significantly enhanced the quality of the semantic representation in language models. As compared to traditional machine learning models, the pre-trained models have achieved impressive results. In this paper, the pre-trained language models are utilized to determine the indexes and sentiment levels of the attribute–viewpoint phrases. Assume that $\mathrm{CSF} = \{\mathrm{CSF}_1, \mathrm{CSF}_2, \ldots \mathrm{CSF}_{12}\}$ represents 12 indexes. $\mathrm{SET} = \{\mathrm{SET}_1, \mathrm{SET}_2, \mathrm{SET}_3\}$ represents three sentiment levels (positive, neutral, negative). Multiply CSF by SET using Cartesian multiplication to list all possible combinations of indexes and sentiment levels.

$$\mathrm{CSF} \times \mathrm{SET} = \left\{ \left(\mathrm{CSF_i}, \mathrm{SET_j}\right) | \mathrm{CSF_i} \in \mathrm{CSF}, \mathrm{SET_j} \in \mathrm{SET} \right\} \tag{3}$$

Therefore, 36 index–sentiment classes have been created [34]. The inputs of models are constructed in three steps, as follows.

Step 1: Concatenate the attribute–viewpoint phrases and reviews to form sentence pairs.

Step 2: Add *[CLS]* and *[SEP]* to the sentence pairs to construct input tokens.

Step 3: Transform the input tokens into input ID according to the lexicon, and construct token type ID to indicate the segments of sentences.

Finally, the inputs of the models consist of input ID and token type ID. The classes of the sentence embedding (h_1) are determined by a fully connected layer. The losses are calculated by Softmax and Cross-Entropy function. The training process for classification based on the pre-trained models is shown in Figure 2.

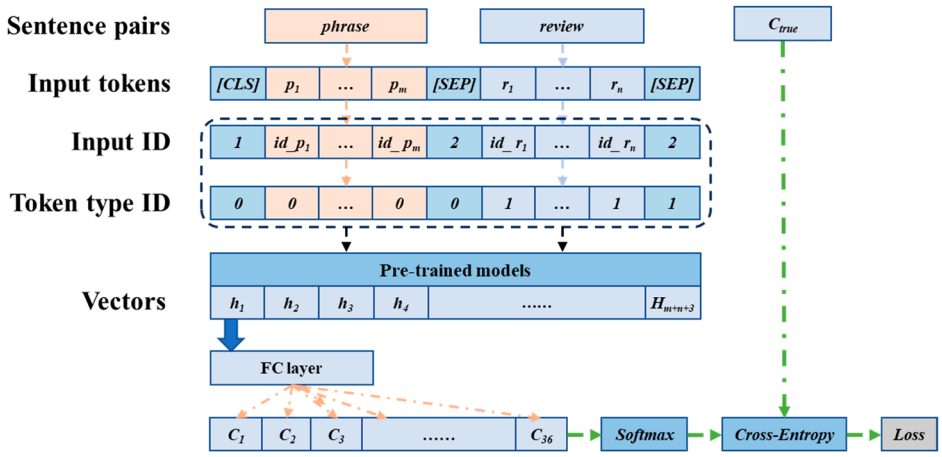


Figure 2. The training process for classification based on the pre-trained models.

In the training phase, a tuple (*phrase*, *review*) is given, and a token sequence is constructed in a form similar to "*[CLS]*, p_1, ⋯, p_m, *[SEP]*, r_1, ⋯, r_n, *[SEP]*", where p_1, ⋯, p_m are words that form a *phrase*, and r_1, ⋯, r_n are words that form a *review*. Moreover, *[CLS]* and *[SEP]* are special tokens in the pre-trained models. The token sequence is subsequently input into the pre-trained models, and the vector of *[CLS]* h_1 is used for classification via an FC layer.

Using the self-attention mechanism, h_1 is derived by a weighted summing-up of the embeddings of all other tokens. Therefore, h_1 aggregates the information for the entire sentence pairs, and can be used for sentence classification. There are alternative methods for acquiring the ultimate classification vector, like averaging for all the embeddings. However, those methods tend to lose important information of h_1, leading to poor classification performance.

3.3. Ranking of Perceived Quality Indexes

PLTS is constructed based on the classification results, and the sentiment levels are treated as discrete random variables. Then, stochastic dominance criteria are applied to establish the stochastic dominance relation matrix concerning each index. Based on the stochastic dominance relation matrix, the stochastic dominance degree matrix of indexes is constructed, and the overall stochastic dominance degree (comprehensive flow) of each index is obtained using the PROMETHEE-II method [35].

3.3.1. Construction of Stochastic Dominance Relation Matrix

Transform the distributions of sentiment levels concerning the indexes in PLTS into discrete percentage distributions. Afterwards, the stochastic dominance criteria are introduced to create the stochastic dominance relation matrix. The specific procedure is as follows.

For all reviews of a certain brand, the amount of reviews is counted by indexes, and the amounts of reviews with certain sentiment levels are counted separately [8,36–38], defined as:

$$Pro_k^{\varepsilon} = \frac{P_k^{\varepsilon}}{P_k},\ k = 1, 2, 3 \cdots 12; \varepsilon = 1, 2, 3 \tag{4}$$

where P_k denotes the amount of phrases of the k-th index, and P_k^{ε} denotes the amount of phrases with the ε-th sentiment level of the k-th index. The cumulative distribution functions of the sentiment levels concerning CSF_k can be obtained:

$$F_k(t) = \sum_{t \geq T^{\varepsilon}} Pro_k^{\varepsilon} \tag{5}$$

Let $E = (e_k)_{12\times 1}$ be the expectation of the cumulative distribution function, where e_k can be expressed by:

$$\mathrm{e}_k = \sum_{\varepsilon=1}^{3} Pro_k^{\varepsilon} T^{\varepsilon} (k = 1, 2, \cdots, 12; \varepsilon = 1, 2, 3) \tag{6}$$

T^{ε} is set to the same value as ε in this paper. On an interval [a,b], if two cumulative distribution functions $F(x) \neq G(y)$ satisfy:

$$H_1(x) = F(x) - G(x) \leq 0 \tag{7}$$

then $F(x)$ dominates $G(y)$ in the first order, denoted as $F(x)$ FSD $G(y)$. Otherwise, if:

$$H_2(x) = \int_a^x H_1(y)dy \leq 0 \tag{8}$$

then $F(x)$ dominates $G(y)$ in the second order, denoted as $F(x)$ SSD $G(y)$. Otherwise, if:

$$H_3(x) = \int_a^x H_2(y)dy \leq 0 \tag{9}$$

$$E(x) > E(y) \tag{10}$$

then $F(x)$ dominates $G(y)$ in the third order, denoted as $F(x)$ TSD $G(y)$, otherwise denoted as '--'.

3.3.2. Construction of Stochastic Dominance Degree Matrix

Based on the stochastic dominance relation matrix, the stochastic dominance degree matrix SUR = [sur(CSF_k, CSF_h)] is constructed using the PROMETHEE-II method by Equation (11).

$$\mathrm{sur}(\mathrm{CSF}_k, \mathrm{CSF}_h) = \begin{cases} 1, r_{kh} \in SD, e_k \geq e_h + a \\ \frac{e_k - e_h}{a}, r_{kh} \in SD, e_h \leq e_k \leq e_h + a \\ --, \text{otherwise} \end{cases} \tag{11}$$

In the equation above, SD = {FSD, SSD, TSD}, and a is the preference threshold:

$$a = \frac{2}{12 \times (12-1)} \sum_{\mathrm{k}=1}^{12} \sum_{\mathrm{h}=1 h \neq k}^{12} b_{kh} \tag{12}$$

$$b_{kh} = \begin{cases} e_k - e_h, e_k \geq e_h \\ 0, e_k < e_h \end{cases} \tag{13}$$

Compute leaving flow ϕ^+, entering flow ϕ^-, and the comprehensive flow ϕ for all indexes. For a certain index CSF_k:

$$\phi^+(\mathrm{CSF}_k) = \sum_{i=1}^{12} \mathrm{SUR}_{k,i} \tag{14}$$

$$\phi^-(\mathrm{CSF}_k) = \sum_{i=1}^{12} \mathrm{SUR}_{i,k} \tag{15}$$

$$\phi(\mathrm{CSF}_k) = \phi^+(\mathrm{CSF}_k) - \phi^-(\mathrm{CSF}_k) \tag{16}$$

The leaving flow $\phi^+(\mathrm{CSF}_k)$ represents the degree to which the index CSF_k is superior to other indexes, the entering flow $\phi^-(\mathrm{CSF}_k)$ represents the degree to which the index CSF_k is inferior to other indexes, and the comprehensive flow $\phi(\mathrm{CSF}_k)$ represents the degree to which the index is overall superior to all other indexes. The greater the comprehensive flow, the better the performance of the index [35].

4. Experiments and Discussions

4.1. Data Collection and Pre-Processing

For effective access to reviews, a web crawler was developed in advance. Then, following the default review display order, reviews from four brands (ALBD, AND, BNL, QPL) on JD Mall were collected in June 2023 using the web crawler; 10,000 reviews from 2020 to 2023 were collected for each brand. The annual distribution of the collected online reviews is illustrated in Figure 3.

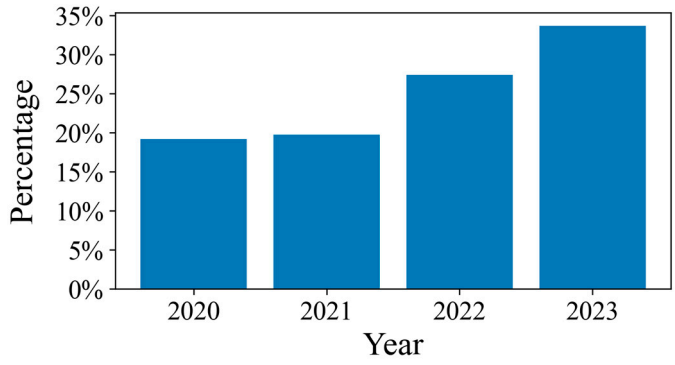


Figure 3. Annual distribution of the collected online reviews.

Considering the concentration of terms, 4000 reviews were selected randomly, and 3918 terms were obtained by Chinese word segmentation. The quantities of the 1000 most frequent terms are displayed in Figure 4.

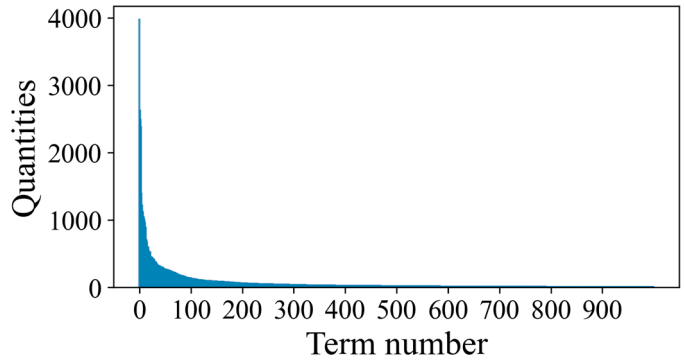


Figure 4. Quantities of the 1000 most frequent terms.

Finally, 1899 nouns and their BERT embeddings were obtained.

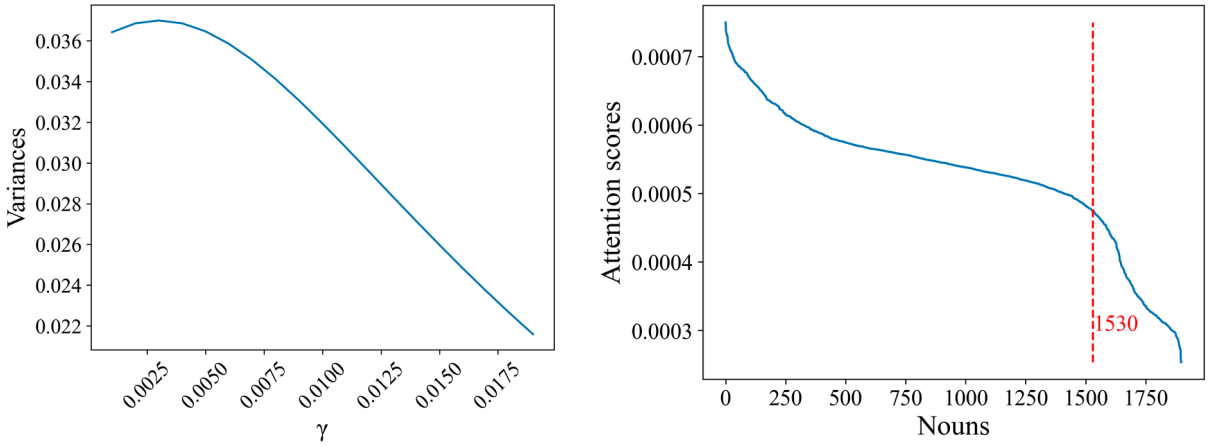
4.2. Extraction of Attribute-Viewpoint Phrases

Based on the BERT embeddings, the attention scores were calculated using the method introduced in Section 3.1. To determine the value of γ in Equation (1), the variances of

standardized attention scores for each noun with varying γ were calculated. The formula of standardized processing is:

$$x' = \frac{x - X_{min}}{X_{max} - X_{min}} \tag{17}$$

where x is the attention score of a noun. The results of variances are shown in Figure 5a. The variance of the attention scores is the highest when γ is set to 0.003, indicating that some nouns are more likely to serve as aspect words than the others. The specific values of the attention scores for each noun are shown in Figure 5b.



(**a**) (**b**)

Figure 5. (**a**): Variances of standardized attention scores with varying γ; (**b**) The attention scores of each noun (γ = 0.003).

According to the principle of the elbow method, 1530 nouns were selected to construct a lexicon of aspect words. The top six words in the attention scores are shown in Table 3.

Table 3. The top six words in the attention scores.

Aspect Words	Attention Scores	Aspect Words	Attention Scores
quality	0.00074981933	necessities	0.00073796633
time	0.00073314965	health	0.00073012454
feeling	0.00071956177	service	0.00070828611

Sentiment words were extracted with regard to the aspect words based on the guidelines specified in Table 2. Then, the aspect words and corresponding sentiment words were merged to form attribute–viewpoint phrases. Examples of sentiment words extraction are shown in Table 4. The amounts of attribute–viewpoint phrases for each brand are listed in Table 5.

Table 4. Examples of sentiment words extraction.

Raw Reviews	Aspect Words	Sentiment Words
The quality is very good, the pattern is nice, and I am also very satisfied with the service. I will come again next time.	quality pattern service	very good nice very satisfied

Table 5. Amounts of attribute–viewpoint phrases for each brand.

Brands	Amounts of Attribute-Viewpoint Phrases
ALBD	20,629
AND	15,336
BNL	17,608
QPL	20,510

4.3. Classification of Reviews

4.3.1. Model Training

Attribute–viewpoint phrases and index-sentiment classes were manually annotated for model training; 1630 reviews were annotated for the training set, and 572 reviews for the test set. Three more pre-trained language models were trained and tested to select the most appropriate one for classification. Ernie 3.0: a unified framework for pre-training large-scale knowledge enhanced models, which can be easily tailored for natural language understanding and generation tasks with zero-shot learning [39]; Roberta: an improved version of BERT, in the training process of Roberta, Byte-Pair Encoding and Dynamic Mask are applied, Next Sentence Pair task is removed [40]; Macbert: a pre-trained model which improves upon Roberta in several ways, including the masking strategy that adopts *MLM* as a correction (Mac) [41].

In the training process, the hyper-parameters were established with the following values: **Epochs**: 20, representing the times the entire training dataset had been traversed. **Batch_size**: 16 or 32, representing the size of each input data provided to the model. **Lr**: 1×10^{-5} or 1×10^{-4}, representing the step size of the network weight adjustment along the gradient direction. Experiments were conducted for varying hyper-parameter configurations. The losses of the training process for each model are displayed in Figure 6.

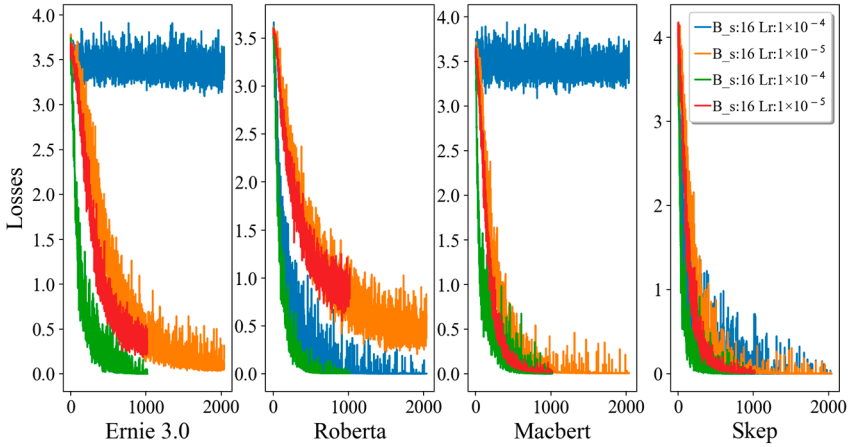


Figure 6. The losses of the training process for each model.

It can be seen that Skep exhibits better loss convergence, implying more effective fitting with the dataset in this paper. Macbert and Ernie 3.0 did not converge at a Batch_size of 16 and Lr of 1×10^{-4}. When Lr was set to 1×10^{-5}, ERNIE 3.0 and Roberta exhibited a tendency to enter local optima, and additional training was required to minimize the losses. Moreover, it was found that utilizing a Batch_size of 32 resulted in more stable losses and smaller fluctuations for each model, thus promoting the models' convergence.

Macro-P, Macro-R, and Macro-F1 were employed as metrics to evaluate the generalization effect of the models. The metrics above were computed using Equations (18)–(20).

$$\text{Macro-P} = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{18}$$

$$\text{Macro-R} = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{19}$$

$$\text{Macro-F1} = \frac{1}{n}\sum_{i=1}^{n} F1_i = \frac{1}{n}\sum_{i=1}^{n} \frac{2P_i R_i}{P_i + R_i} \tag{20}$$

The optimum Macro-F1 values for each model with varying hyper-parameters are presented in Table 6. Additionally, the results are compared by several traditional models based on word vectors using Tf-idf.

Table 6. Optimum Macro-F1 values for each model with varying hyper-parameters on the test set.

Model	Batch_Size	Lr	Macro-P	Macro-R	Macro-F1
Ernie-3.0	32	1×10^{-4}	0.78	0.76	0.76
Roberta	16	1×10^{-4}	0.74	0.73	0.73
Macbert	16	1×10^{-5}	0.80	0.78	0.77
Skep	32	1×10^{-5}	0.83	0.77	0.77
Text_CNN	32	1×10^{-5}	0.55	0.49	0.50
Tf-idf + SVC	-	-	0.51	0.49	0.46
Tf-idf + SGD	-	-	0.48	0.47	0.46
Tf-idf + RandomForest	-	-	0.40	0.38	0.36

According to Table 6, the traditional models exhibited unsatisfactory performance in this paper, whereas the pre-trained models demonstrated exceptional performance. Among the pre-trained models, Macbert and Skep can achieve almost the same level of effect, but, according to Figure 6, Skep exhibits better loss convergence, so the following classification task is achieved by Skep.

4.3.2. Classification Results

The reviews of the four brands were transformed and classified by Skep following the method described in Section 3.2. The results of the review classification for each brand are shown in Figure 7.

Figure 7 states that the clothing industry possesses numerous significant supplemental features, with a 20% variance in External between the four brands' highest and lowest levels. Moreover, there exist concealed hazards in safeguarding consumer information, with low Security and Certification levels for each brand.

The distribution of reviews related to each index for each brand is shown in Figure 8.

For the overall perceived quality of clothing brands, reviews primarily relate to products rather than services, indicating that the attributes of products are the main influencing factors in forming perceived quality, while services have a relatively supplementary role.

The cumulative percentage of the first six indexes attained 90%, and Comfort is the most valued index by consumers. Consequently, clothing brands should focus on Comfort, External, Protection, Fineness, Professionalism, and Appearance indexes.

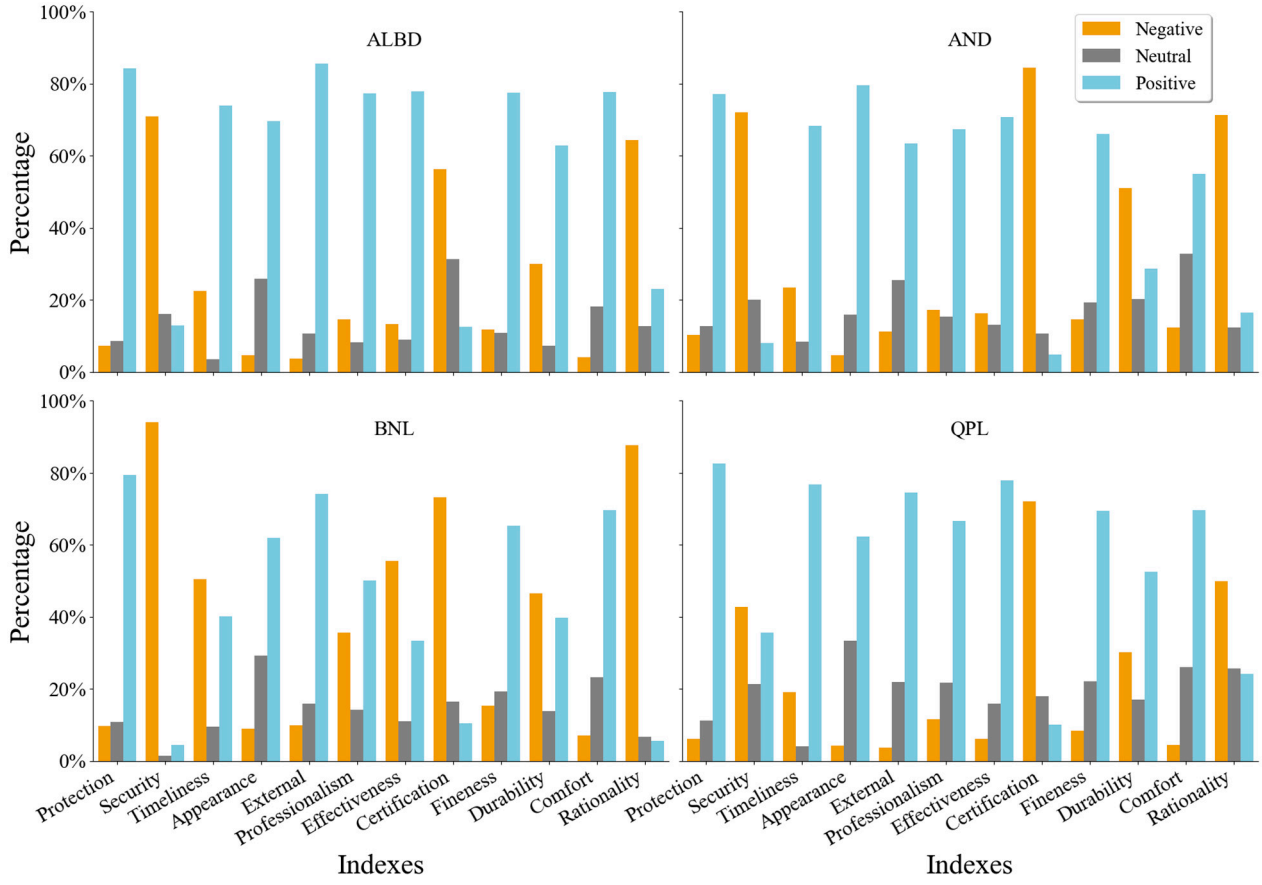


Figure 7. Results of review classification for each brand.

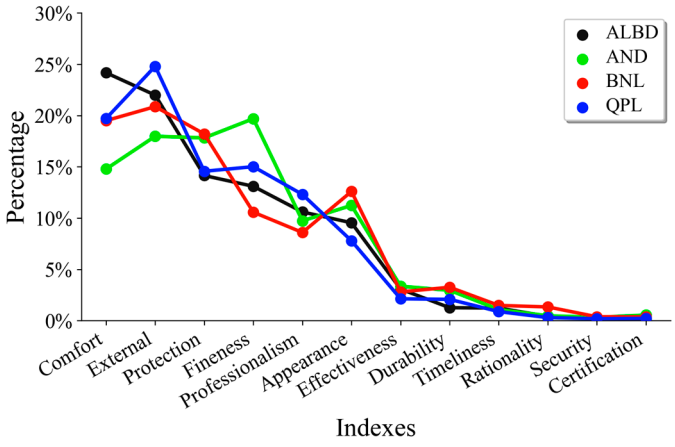


Figure 8. The distribution of reviews related to each index for each brand.

4.4. Ranking of Perceived Quality Indexes for Clothing Brands

After classification, discrete percentage distributions of sentiment levels for each index were obtained, and the PLTS were constructed. Then, stochastic dominance criteria and the PROMETHEE-II method were used to compute the comprehensive flow of the chosen brands. The result is shown in Table 7.

Table 7. Comprehensive flow of the selected brands.

Indexes	ALBD	AND	BNL	QPL
Comfort	4.6	2.9	7	4.2
Durability	−2	−5.9	−1.9	−4.2
Fineness	3.3	3.5	5.5	3.2
Protection	4.8	5.6	7.2	5.2
Certification	−8.8	−9.6	−7.5	−10.8
Appearance	3.4	6.6	6.3	3.5
External	5.9	3.7	6.8	5.2
Timeliness of response	1	2.7	−2.3	2.6
Effectiveness of communication	3.1	3.7	−3.5	4.8
Security of information	−9.6	−8.6	−9.5	−7.2
Rationality of compensation	−8.7	−8	−9	−8.6
Professionalism of service	2.8	3.3	0.9	2.2

Based on the comprehensive flow presented in Table 7, the ranking results of the four brands are shown in Figure 9.

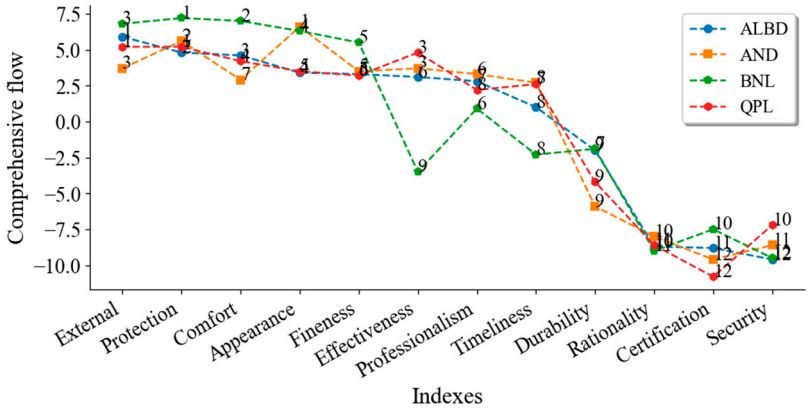


Figure 9. Ranking results of the four brands.

As demonstrated in Figure 9, the ranking results of the four brands exhibit a consistent pattern, and it can be seen that the evaluation of products is superior to that of services, indicating that clothing brands place a higher emphasis on products than services.

The differences displayed in Figure 9 for each brand highlight variations in the primary indexes of each brand. Specifically, BNL exhibits a more substantial discrepancy between its services and products compared to the other three brands, suggesting that BNL concentrates more resources on its products and spends fewer resources on its services than the other brands.

4.5. Comprehensive Analyses and Suggestions

External and Protection are highly ranked across indexes for all the brands and are highly valued by consumers. However, Figure 7 indicates that there are fewer variations between those brands. This phenomenon suggests that clothing consumers tend to have a positive perception of brands due to external and safety-related factors.

Comfort can show the differences between brands to some extent. According to Figure 9, Comfort ranks highly in ALBD, BNL, and QPL, but ranks only seventh in AND, indicating the differences between AND and the other brands. Additionally, Figure 7 reveals many neutral reviews of Comfort within AND, suggesting a need for improvement in product comfort to turn neutral reviews into positive ones.

The rates of Appearance also fluctuate considerably between brands, but Appearance has received limited attention from consumers. According to Figure 8, only 10% of the reviews are related to Appearance. The excellent rate of Appearance in AND could not

bring a significant improvement for it, so it is suggested that AND reduce the cost of its appearance design to some extent.

According to Figure 9, BNL's Effectiveness ranks very low, at ninth place, while the effectiveness of other brands ranks higher. Unlike other brands, BNL receives more negative reviews than positive in this index; therefore, BNL needs to focus on enhancing the quality of customer service responses.

In terms of perceived service quality, communication quality (Effectiveness and Timeliness) surpasses delivery quality (Professionalism) for ALBD and AND. Regarding the correlation between Effectiveness and Timeliness, the balance between these two indexes is impacted by the commonly used auto-responder bots, which have their pros and cons. On the one hand, auto-responder bots enable brands to promptly respond to basic consumer demands, potentially enhancing efficiency. However, if the bots prove ineffective in resolving issues and manual customer service fails to respond promptly, it may result in a lackluster response. To improve the perceived service quality of consumers during the communication process, brands should establish flexible auto-responder bots that collaborate rationally with manual customer service agents to resolve issues effectively.

5. Conclusions

From the perspective of consumers, this paper has derived the evaluation index system of perceived quality for clothing brands, which provides a theoretical framework for the perceived quality evaluation of clothing brands. Based on fine-grained sentiment analysis and multi-criteria decision making methods, this paper has proposed a multi-stage model ECRM for perceived quality evaluation. In the first stage, RBF and dependency parsing are used to extract attribute–viewpoint phrases. Next, reviews are classified using pre-trained language models. Finally, indexes are then ranked using stochastic dominance criteria and PROMETHEE-II. ECRM is presented as a system composed of specific components and may serve as a reference for other industries.

Clothing is intimately connected with individuals. The results of empirical experiments show that certain common and distinct features of perceived quality exist among clothing brands. Of all the indexes, clothing brands should give priority to the comfort of clothing products and other factors closely related to people. In the environment of online shopping, consumers' demand for service quality is not particularly significant compared to product quality.

Author Contributions: Conceptualization, J.C. and J.Z.; methodology, J.C.; validation, M.R.; formal analysis, Y.F.; investigation, M.R.; resources, J.Z.; data curation, M.R.; writing—original draft preparation, M.R.; writing—review and editing, Y.F. and J.C.; visualization, M.R.; supervision, Y.F.; project administration, J.C.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National key research and development program, grant number 2019YFB1405300, the Project of Cultivation for Young Top-notch Talents of Beijing Municipal Institutions "Research on the comprehensive quality intelligent service and optimized technology for small medium and micro enterprises" (Grant No. BPHR202203233), and the National Natural Science Foundation of China "Research on the influence and governance strategy of online review manipulation with the perspective of E-commerce ecosystem" (Grant No. 72174018).

Data Availability Statement: The datasets generated during and analyzed during the current study are available from the corresponding author on reasonable request.

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

References

1. China Clothing Association. *2021–2022 China Clothing Industry Development Report*, 1st ed.; China Textile Publishing House: Beijing, China, 2022; pp. 19–21.
2. Chen, X.; Liang, H.; Gao, Y.; Xu, W. A method based on the disappointment almost stochastic dominance degree for the multi-attribute decision making with linguistic distributions. *Inf. Fusion* **2020**, *54*, 10–20. [CrossRef]

3. Li, L.; Yang, W.S.; Xie, Y.Q. Research on Quality Information Asymmetry in the E-commerce Market. *Manag. Rev.* **2004**, *3*, 25–30+63–64.
4. Valarie, A.Z.; Leonard, L.B.; Parasuraman, A. Consumer perceptions of price quality and value: A means end model and synthesis of evidence. *J. Mark.* **1988**, *52*, 35.
5. Chang, Y.P.; Zhu, D.H.; Zhang, J.L. Research on the Relationship between Knowledge Sharing in Virtual Communities and Consumer Brand Conversion. *J. Manag.* **2009**, *6*, 1536–1540+1554.
6. Zhu, Z.Z.; Liu, F.; Wei, H.Y. Research on the Influence Mechanism of Product Design on Consumer Word of mouth Communication—Mesomeric effect of Emotion and Perceived Quality. *J. Dalian Univ. Technol. (Soc. Sci. Ed.)* **2022**, *43*, 55–62.
7. Li, S.; Liu, F.; Zhang, Y.; Zhu, B.; Zhu, H.; Yu, Z. Text Mining of User-Generated Content (UGC) for Business Applications in E-Commerce: A Systematic Review. *Mathematics* **2022**, *10*, 3554. [CrossRef]
8. Yu, S.; Du, Z.; Lin, X.; Luo, H.; Wang, J. A Stochastic Dominance-Based Approach for Hotel Selection under Probabilistic Linguistic Environment. *Mathematics* **2020**, *8*, 1525. [CrossRef]
9. Cao, W.; Yang, X.; Yang, Y. A Large-Scale Reviews-Driven Multi-Criteria Product Ranking Approach Based on User Credibility and Division Mechanism. *Mathematics* **2023**, *11*, 2952. [CrossRef]
10. Li, Y. Research on Customer Perceived Product Quality Evaluation Based on Emotional Analysis. Master's Thesis, Hefei University of Technology, Hefei, China, 2020.
11. He, Y.K. Research on Evaluation of E-commerce Logistics Service Quality Based on Emotional Analysis. Master's Thesis, Nanjing University, Nanjing, China, 2021.
12. Zhou, M.; Li, Y. Customer Perceived Product Quality Evaluation Based on Comments on Big data. *Sci. Technol. Dev.* **2020**, *16*, 804–810.
13. Zheng, S.Y.; Wang, P.; Din, H.; Tan, G.X. A study on the user experience of museum digital services based on aspect-level sentiment analysis. *Intell. Sci.* **2022**, *40*, 171–178.
14. Lu, S.; Liu, M.; Yin, L.; Liu, X.; Zheng, W. The multi-modal fusion in visual question answering: A review of attention mechanisms. *PeerJ Comput. Sci.* **2023**, *9*, 1400. [CrossRef] [PubMed]
15. Zhang, Y.; Shao, Z.; Zhang, J.; Wu, B.; Zhou, L. The effect of image enhancement on influencer's product recommendation effectiveness: The roles of perceived influencer authenticity and post type. *JRIM* **2023**, *4*, 385–401. [CrossRef]
16. Cheng, L.; Yin, F.; Theodoridis, S.; Chatzis, S.; Chang, T.-H. Rethinking Bayesian Learning for Data Analysis: The art of prior and inference in sparsity-aware modeling. *IEEE Signal Process* **2022**, *6*, 22. [CrossRef]
17. Nie, W.; Bao, Y.; Zhao, Y.; Liu, A. Long Dialogue Emotion Detection Based on Commonsense Knowledge Graph Guidance. *IEEE Trans. Multimed.* **2023**, *8*, 1–15. [CrossRef]
18. Liu, X.; Zhou, G.; Kong, M.; Yin, Z.; Yin, L.; Zheng, W. Developing Multi-Labelled Corpus of Twitter Short Texts: A Semi-Automatic Method. *Systems* **2023**, *11*, 390. [CrossRef]
19. Zeng, X.J.; Ye, X.Q.; Liu, D. Research on User Satisfaction Analysis Based on Fine-Grained Viewpoint Mining and Kano Modeling. *Comput. Eng. Sci.* **2023**, *45*, 701–710.
20. Liu, X.X.; Chen, Z.Y. Service quality evaluation and service improvement using online reviews: A framework combining deep learning with a hierarchical service quality model. *Electron. Commer. Res. Appl.* **2022**, *54*, 101174. [CrossRef]
21. Tulkens, S.; Van, C.A. Embarrassingly Simple Unsupervised Aspect Extraction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020.
22. Tian, H.; Gao, C.; Xiao, X.; Liu, H.; He, B. SKEP: Sentinel Knowledge Enhanced Pre training for Sentinel Analysis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020.
23. Feng, K.; Yang, Q.; Chang, X.Y. Fresh food e-commerce customer satisfaction evaluation based on online comments and stochastic dominance criteria. *China Manag. Sci.* **2020**, *29*, 205–216.
24. Majid, B.; Kazemzadeh, R.; Albadvi, A. PROMETHEE: A comprehensive literature review on methodologies and applications. *Eur. J. Oper. Res.* **2010**, *200*, 198–215.
25. Liu, X.F.; Ji, X.F.; Wang, J. Study on Perceived Quality of Silk Clothing. *Silk* **2007**, *3*, 49–52.
26. Valarie, A.Z.; Leonard, L.B.; Parasuraman, A. Communication and Control Processes in the Delivery of Service Quality. *J. Mark.* **1988**, *3*, 35–48.
27. Parasuraman, A.; Valarie, A.Z.; Leonard, L.B. SERVQUAL: A Multiple Item Scale for Measuring Consumer Perceptions of Service Quality. *J. Retail.* **1988**, *64*, 12–40.
28. Zhao, W.H.; Xiong, X.M. Measurement and Management of Service Quality in Online Retail: Based on the Chinese Context. *Manag. Rev.* **2015**, *27*, 120–130.
29. Eugene, F.; Dianna, L.; Dharuv, G. Development of a multidimensional measure of perceived product quality. *J. Qual. Manag.* **1997**, *2*, 87–111.
30. William, B.D.; Kent, B.M.; Dharuv, G. Effects of Price, Brand, and Store Information on Buyers' Product Evaluations. *J. Mark. Res.* **1991**, *28*, 307.
31. Jacob, D.; Chang, M.W.; Lee, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), Minneapolis, MN, USA, 2–7 June 2019.

32. Zhang, J.; Lu, X.; Liu, D. Deriving customer preferences for hotels based on aspect level sentient analysis of online reviews. *Electron. Commer. Res. Appl.* **2021**, *49*, 94–101. [CrossRef]
33. Sun, B.S.; Ao, C.L.; Wang, J.X. Research on satisfaction evaluation of ecotourism based on online Text mining. *Oper. Res. Manag.* **2022**, *31*, 165–172.
34. Wan, H.; Yang, Y.F.; Du, J.F.; Liu, Y.N.; Qi, K.X.; Pan, J.Z. Target-Aspect-Sentiment Joint Detection for Aspect-Based Sentiment Analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 7–12 February 2020.
35. Fan, Z.P.; Xi, Y.; Liu, Y. Supporting consumer's purchase decision: A method for ranking products based on online multi-attribute product ratings. *Soft Comput.* **2018**, *22*, 5247–5261. [CrossRef]
36. Zhang, Y.; Fan, Z.P. A stochastic multi-attribute decision-making method based on stochastic dominance. *J. Syst. Manag.* **2010**, *19*, 371–378.
37. Kazimierz, Z.; Jean, M.M. Multiattribute Analysis Based on Stochastic Dominance. *Models Exp. Risk Ration.* **1994**, *20*, 225–248.
38. Samuel, B.G.; Jeffrey, L.R. Probabilistic domain criteria for comparing unknown alternatives: A tutorial. *Omega* **2009**, *37*, 346–357.
39. Sun, Y.; Wang, S.; Feng, S. ERNIE 3.0: Large-Scale Knowledge Enhanced Pre-Training for Language Understanding and Generation. Available online: https://arxiv.org/abs/2107.02137 (accessed on 30 August 2023).
40. Liu, Y.; Ott, M.; Goyal, N. RoBERTa: A Robustly Optimized BERT Pretraining Approach. Available online: https://arxiv.org/abs/1907.11692 (accessed on 30 August 2023).
41. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Wang, S.; Hu, G. Revisiting Pre-Trained Models for Chinese Natural Language Processing. In *Findings of the Association for Computational Linguistics: EMNLP*; Association for Computational Linguistics: Abu Dhabi, United Arab Emirates, 2022.

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.