Asia Pacific Journal of Information Systems
Vol. 34 No. 2 (June 2024), 372-399

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2024.34.2.372

Investigating the Impact of Discrete Emotions Using Transfer Learning Models for Emotion Analysis: A Case Study of TripAdvisor Reviews

Dahee Lee[a], Jong Woo Kim[b,*]

[a] *Ph.D. Candidate, School of Business, Hanyang University, Korea*
[b] *Professor, School of Business, Hanyang University, Korea*

A B S T R A C T

Online reviews play a significant role in consumer purchase decisions on e-commerce platforms. To address information overload in the context of online reviews, factors that drive review helpfulness have received considerable attention from scholars and practitioners. The purpose of this study is to explore the differential effects of discrete emotions (anger, disgust, fear, joy, sadness, and surprise) on perceived review helpfulness, drawing on cognitive appraisal theory of emotion and expectation-confirmation theory. Emotions embedded in 56,157 hotel reviews collected from TripAdvisor.com were extracted based on a transfer learning model to measure emotion variables as an alternative to dictionary-based methods adopted in previous research. We found that anger and fear have positive impacts on review helpfulness, while disgust and joy exert negative impacts. Moreover, hotel star-classification significantly moderates the relationships between several emotions (disgust, fear, and joy) and perceived review helpfulness. Our results extend the understanding of review assessment and have managerial implications for hotel managers and e-commerce vendors.

Keywords: Review Helpfulness, Discrete Emotions, Text Mining, Emotion Analysis, Hotel Star-Classification

Ⅰ. Introduction

Online reviews have become essential for decision making by consumers. According to a consumer survey conducted by BrightLocal (2022), 77% of consumers always or regularly read online reviews before making purchase decisions in 2021. Information overload occurs as the volume of reviews increases (Mudambi and Schuff, 2010; Park and Lee, 2008). To enable consumers to efficiently identify reviews that are worth reading, a number of e-commerce platforms have adopted voting systems through which consumers can evaluate the helpfulness of reviews. Helpful votes enable consumers to better

*Corresponding Author. E-mail: kjw@hanyang.ac.kr

understand the quality and performance of a product or service, influencing their attitudes and purchase decisions (Jiang and Benbasat, 2007; Xie et al., 2011).

Exploring predictors of review helpfulness has continuously gained attention from academic researchers and practitioners. Previous studies have revealed that review helpfulness can be affected by characteristics of reviews, reviewers, and products/services, such as review emotions (Ahmad and Laroche, 2015; Ren and Hong, 2019; Wang et al., 2019), reviewer expertise (Filieri et al., 2018; Ghose and Ineirotis, 2010; Weiss et al., 2008), and product type (Mudambi and Schuff, 2010; Pan and Zhang, 2011; Racherla and Friske, 2012). In the travel and tourism industry, several hotel-related variables such as photos of the hotel, the number of rooms in the hotel, and the location were further investigated (Filieri et al., 2021; Filieri et al., 2018; Rhee and Yang, 2015).

Recent studies focused on the influence of feelings or emotional states included in reviews by measuring review sentiment, which is featured with one-dimensional valence (i.e., polarity: positive vs. negative) (Baek et al., 2012; Chua and Banerjee, 2016; Lee et al., 2017; Pan and Zhang, 2011; Peng et al., 2014; Yang et al., 2020). In particular, most of the prior studies in the hotel industry have investigated the sentiment scores of hotel attributes such as location, meal, and service, without relating them to review helpfulness (Chen et al., 2022; Khotimah and Sarno, 2019). However, such valence-based approaches have been challenged since they not appropriate for discovering the effects of multi-dimensional emotions embedded in reviews. People-to-people service experience may have impact on consumers' different emotions. For instance, the consumer might feel happy due to the service recovery whereas the other might feel angry and even disgust when the service recovery fails. The emotional cues in reviews could transfer to the mood or attitude of a reader, resulting in evaluating review helpfulness (Ludwig et al., 2013; Wang et al., 2019). Thus, examining the impacts of multiple emotions on decision making has been an emerging trend (Ahmad and Laroche, 2015; Shah and Lee, 2022; Wang et al., 2019). To the best of our knowledge, there is a lack of research comprehensively addressing the differential effects of positive and negative emotions. Furthermore, the methods used to identify emotions in online reviews in the previous studies were primarily dictionary-based, and therefore unable to detect negative emotions. Such methods also suffer from out-to-date vocabulary coverage. With the help of advanced text mining techniques, it has become possible to identify more features embedded in review texts (Fan, 2021). Therefore, we suggest a novel method to measure emotions by applying a transfer learning model for emotion analysis. This model can capture emotions from reviews even though they are implicitly expressed without using specific emotional words.

In this study, we investigate the differential effects of six discrete positive and negative emotions (i.e., anger, disgust, fear, joy, sadness, and surprise) on review helpfulness based on Ekman's basic emotion theory (1992). By constructing machine learning based emotion classifiers, six emotions are extracted from reviews and used when measuring discrete emotion variables as predictors of review helpfulness. Second, we predict that the impacts of specific emotions are moderated by hotel classification, referring to previous findings that hotel classification affects consumer assessments (Filieri et al., 2018; Filieri et al., 2021; Silva, 2015). Moderating effects regarding relationships between specific emotions and review helpfulness remain unexplored (Filieri et al., 2018; Karimi and Wang, 2017). Thus, this study may aid understanding the inconclusive results of previous research.

Our study makes several contributions to the literature. First, we devise a new method to measure discrete emotions embedded in reviews by employing a transfer learning model. We could capture contextually expressed emotions more precisely, which is impossible through the traditional dictionary-based approach. Second, we provide a new insight by validating the effect of the hotel star-classification as a moderator in the relationship between emotional content and review helpfulness. Third, this study extends the current understandings of the nuance between discrete emotions in review evaluations. The results of this study may allow hotel managers and online vendors to readily identify helpful reviews.

The rest of the paper is organized as follows. Relevant studies are described in Section 2, focusing on review helpfulness, theories of emotions, and hotel classification. Section 3 presents the research model and our research hypotheses. Section 4 describes our methodology including data collection, data processing introducing a new emotion assessment method, variable operationalization, and statistical analysis. In Section 5, we report our empirical results. Conclusions regarding findings, limitations, and future research directions are presented in Section 6.

Ⅱ. Conceptual Background

2.1. Discrete Emotions

Emotions refer to valenced feeling states representing evaluative reactions to events, agents or objects (Kim and Gupta, 2012). To define emotions more precisely, Ekman (1992) stated that six basic emotions (i.e., anger, disgust, fear, joy, sadness, and surprise) can be experienced separately but are universally recognized, and constitute fundamental elements that combine to form compound emotions. We chose these six emotions as the subject of this study, as the origin of the theory is firmly grounded and widely accepted by psychological studies. The theory also takes into account various negative emotions as well as positive emotions. Several studies have concentrated on negative emotional content (Craciun et al., 2020; Li et al., 2020; Ren and Hong, 2019; Yin et al., 2014) because negative information usually plays a greater role than positive information in consumer decision making, as suggested by the negativity bias effect (Rozin and Royzman, 2001).

Westbrook and Oliver (1991) further defined consumption emotions as emotional responses elicited during shopping experiences, and theories of emotions began to be widely used to interpret message senders' feelings and receivers' responses in information systems and marketing studies (Craciun et al., 2020; Ren and Hong, 2019). According to the cognitive appraisal theory of emotions (CAT), distinct emotions can be differentiated based on several appraisal dimensions (e.g., valence, certainty, personal/situational control) that relate to events that evoked the corresponding emotion (Lazarus, 1991; Smith and Ellsworth, 1985). These dimensions were discovered to be closely associated with discrete consumption emotions (Lerner and Keltner, 2001). For instance, certainty and control are key dimensions that separate anger from fear, even though they belong to the same valence (negative) of emotions. Anger is associated with certainty about events and individual control for negative events, while fear is characterized by uncertainty and situational control.

Several scholars have proved that same valence emotions can have contrary impacts on review helpfulness while different valence emotions may have the same effects, based on the cognitive appraisal theory of emotions. Yin et al. (2014) explored the differential effects of two negative emotions, anger and anxiety, and found that anxiety-embedded reviews are perceived to be more helpful than anger-embedded reviews. Ren and Hong (2019) studied three negative emotions (i.e., anger, fear, and sadness) and confirmed that fear has a positive effect upon review helpfulness while anger and sadness have negative effects.

As of yet, only a few studies have comprehensively addressed the effects of both positive and negative emotions (Ahmad and Laroche, 2015; Shah and Lee, 2022; Wang et al., 2019). These studies demonstrated contrasting findings, possibly due to differences in specific conditions and contexts of reviews (Pan and Zhang, 2011) since the studies examined reviews of kitchen appliances (Ahmad and Laroche, 2015), restaurants (Wang et al., 2019), and physicians (Shah and Lee, 2022). Our study contributes to this research field by examining how positive and negative emotions affect review helpfulness in the context of hotel reviews.

For the extraction of emotions embedded in reviews, prior studies (Ren and Hong, 2019; Shah and Lee, 2022; Wang et al., 2019; Yin et al., 2014) used a dictionary-based approach. In this approach, researchers usually count numbers of emotional words based on predefined dictionaries such as the National Research Council (NRC) emotion lexicon (Mohammad and Turney, 2010). The approach may not appropriately capture emotions in some cases because it cannot detect negative emotions. For example, although the sentence “I am not satisfied” expresses a negative emotion, the dictionary-based approach will recognize the positive word “satisfied” and be unable to capture the meaning of “not”. In contrast, in this study we newly propose a transfer learning-based approach to extract emotions that can detect emotions not only explicitly but implicitly embedded in the context of reviews by generating context-dependent word embeddings, addressing the limitations of dictionary-based methods.

2.2. Review Helpfulness

Review helpfulness is defined as the subjective assessment of review quality by others, which is adopted as perceived diagnosticity (Cao et al., 2011; Mudambi and Schuff, 2010). To address the issue of information overload arising in online reviews (Filieri et al., 2021; Mudambi and Schuff, 2010), many online platforms (e.g., TripAdvisor.com, Amazon.com, Yelp.com) assess the perceived value of reviews by allowing consumers to vote on helpfulness. “Helpful” votes are expected to support the consumer decision-making process, affect purchase intentions, and enhance the value of business platforms (Filieri, 2015; Otterbacher, 2009; Wang et al., 2019).

A growing number of studies in various domains including the travel and tourism industry have investigated the antecedents of review helpfulness over the past two decades. They have discovered that a variety of review, reviewer, and product/service characteristics affect review helpfulness, emphasizing that it is a multi-faceted concept (Huang et al., 2015; Wang et al., 2019). Regarding review features, researchers have studied review rating/extremity (Filieri et al., 2018; Liu and Park, 2015; Mudambi and Schuff, 2010; Ren and Hong, 2019; Zhou et al., 2020), length (Baek et al., 2012; Huang et al., 2015; Liu and Park, 2015; Mudambi and Schuff, 2010; Ren and Hong, 2019), valence/sentiment (Baek et al., 2012; Chua and Banerjee, 2016; Lee et al., 2017; Pan and Zhang, 2011; Peng et al., 2014; Racherla and Friske, 2012; Salehan and Kim, 2016; Yang et al., 2021), emotions (Ahmad and Laroche, 2015; Ren and Hong, 2019; Shah and Lee, 2022; Wang et al., 2019; Yin et al., 2014), and photos (Filieri et al., 2021; Filieri et al., 2018; Lin et al., 2012; Ma et al., 2018). Reviewer features involve reviewer expertise (Filieri et al., 2018; Ghose and Ipeirotis, 2010; Lee et al., 2017; Weiss et al., 2008), gender (Craciun and Moore, 2019), and identity disclosure (Liu and Park, 2015; Racherla and Friske, 2012). With regard to product/service characteristics, product type (Pan and Zhang, 2011; Racherla and Friske, 2012; Mudambi and Schuff, 2010), price (Baek et al., 2012; Zhu et al., 2014), and total review count (Baek et al., 2012; Craciun et al., 2020; Pan and Zhang, 2011; Wang et al., 2019; Zhu et al., 2014) have been studied.

Most existing studies, which explored emotions embedded in the online reviews, adopted a simple positive-negative continuum to map emotions (Baek et al., 2012; Chua and Banerjee, 2016; Garcia and Schweitzer, 2011; Lee et al., 2017; Pan and Zhang, 2011; Peng et al., 2014; Yin et al., 2017; Zhou and Guo, 2017). However, such valence-based approach to measuring emotions has been criticized for its lack of ability to catch nuances between emotions that differ little in terms of valence (Fan, 2021; Fontaine et al., 2007; Yin et al., 2014). To capture the subtleties of emotions in order to truly understand consumer reviews, there is a growing emphasis on the study of the differential effects of discrete emotions (Craciun et al., 2020; Fan, 2021). Recent advances in natural language processing and text mining techniques allow researchers to extract numerous emotional dimensions. This study contributes to this line of research by analyzing the relationship between review helpfulness and specific emotions extracted from reviews using transfer learning methods.

2.3. Hotel Star-classification

Hotel classification sorts a variety of accommodations into categories using stars, crowns, or flowers (Callan, 1998). It is used to support consumers and to raise or lower their expectations of hotel attributes (Rhee and Yang, 2015). Each country employs different classification criteria. For instance, the star ratings in many European countries are managed by designated government agencies or private organizations. The U.S. maintains various classification systems, with the most highly recognized assessments among being Forbes travel guide hotel ratings and American Automobile Association's Diamond ratings. This study is based on the Forbes and AAA systems since they are used by TripAdvisor, a website that is the source of our experimental data.

Prior studies of determinants of review helpfulness have explored the moderating role of product type, i.e., experience vs. search (Chua and Banerjee, 2016; Mudambi and Schuff, 2010; Racherla and Friske, 2012; Ren and Hong, 2019). Shah and Lee (2022) demonstrated that disease severity moderates the associations between positive or negative emotions and perceived review helpfulness. However, research efforts are still necessary to discover moderators affecting the relationships between emotions and review helpfulness. The current research focuses on the moderating effect of hotel star-classification, which has been shown to be influential on consumer assessment (Filieri et al., 2021; Filieri et al., 2018; Silva, 2015). We explored how the impacts of discrete emotions on review helpfulness vary depending on the hotel star-rating, i.e., high vs. low, which has not been addressed by earlier research.

2.4. Emotion Analysis

Emotion analysis aims to identify multiple emotional dimensions like anger, fear, joy, and sadness specifically through content expression, which is different from sentiment analysis where only positive, neutral, or negative feelings are detected (Hu et al., 2021). As stated in Section 2.2, unlike sentiment analysis, emotion analysis can recognize the nuances between a variety of emotions even when they are slightly different from each other with regard to valence (Fan, 2021; Fontaine et al., 2007; Yin et al., 2014).

There are several methods of emotion analysis. First, dictionary-based or lexicon-based methods use predefined dictionaries of emotional words, such as the NRC emotion lexicon (Mohammad and Turney, 2010) or LIWC dictionary (Pennebaker et al., 2015), and matches words with data to detect emotions. Most previous literature on the effect of emotions on perceived review helpfulness adopted these methods to measure emotion variables (Ren and Hong, 2019; Shah and Lee, 2022; Wang et al., 2019; Yin et al., 2014). However, this approach has trouble dealing with negators that reverse contextual meaning (Hardeniya and Borikar, 2016). For example, in the case "the staff were not friendly at all," only the joyful emotion would be detected through the dictionary-based method due to presence of the word "friendly," ignoring the meaning of the negator "not." Furthermore, dictionary-based methods suffer from the problem of word sense disambiguation, which refers to the process of distinguishing syntactically or semantically different meanings of a word (Hardeniya and Borikar, 2016). For instance, although the word "love" has various senses like "feeling of affection," "person someone loves," or "score of zero in tennis," such methods consider the word "love" to have the same meaning in every text.

On the other hand, machine learning methods depend on algorithms derived from training data that utilize linguistic features derived from texts (Provost and Kohavi, 1998), thus having the ability to overcome problems of dictionary-based approaches. Recent machine learning research focuses on deep learning algorithms such as Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) algorithms, which are based on artificial neural networks that are fed with input data and automatically find word representation necessary for a corresponding task, which in our study is emotion analysis (Janiesch et al., 2021). No prior studies on the impact of emotions on review helpfulness have attempted to operationalize emotion variables with machine learning or deep learning approaches.

In this study we employed transfer learning models, which are considered the most advanced architecture used in deep learning models today. Transfer learning achieves performance improvement by transferring knowledge learned from the source domain to the target domain (Tan et al., 2018). We specifically applied three learning models - RoBERTa (a Robustly Optimized BERT pre-training Approach; Liu et al., 2019), DistilBert (a distilled version of BERT; Sanh et al., 2019), ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately; Clark et al., 2020), each of which was pre-trained on a large amount of different data to generate word embeddings. Several emotion classifiers were constructed based on those learning models by training them on the corpus. We then compared their performances to find the emotion classifier offering the best performance. The classifier of the best performance was applied to the review texts to detect emotions explicitly or implicitly embedded in the review context. Finally, the emotions were used as independent variables of our research model to be statistically analyzed.

Ⅲ. Research Model and Hypotheses

The main objective of this research is to explore how positive and negative emotions influence perceived review helpfulness. As summarized in Section 2.2, CAT indicates that six discrete emotions (i.e., anger, disgust, fear, joy, sadness, and surprise) may have dissimilar impacts on information processing and decision making, as they are associated with different appraisal dimensions (Lazarus, 1991; Smith and Ellsworth, 1985). According to CAT, anger is characterized by high certainty and great individual/situational control (Smith and Ellsworth, 1985). Information with higher certainty and confidence is recognized as useful than more ambiguous information (Sniezek and Van Swol, 2001). Furthermore, anger promotes deeper information processing and angry individuals dedicate more effort to reading messages (Nabi, 2002). Previous research has also found that anger-embedded reviews have positive impacts on readers' assessments of review helpfulness in the context of consuming experience goods (Li et al., 2020; Wang et al., 2019). Consequently, we assume that anger has a positive impact on review helpfulness.

Fear and surprise, in situations where other shopping options are available, facilitate systematic information processing when perceiving high risk in CAT (Ferrer et al., 2016; Lazarus, 1991; Smith and Ellsworth, 1985). Especially, fear-embedded messages are likely to evoke fear and facilitate readers to accept opinions which emphasize the possible detrimental consequences (Dillard et al., 1996). Consistent with previous findings (Shah and Lee, 2022; Wang et al., 2019), we hypothesize that the impacts of fear and surprise on perceived review helpfulness are positive.

In contrast with fear and surprise, disgust is associated with heuristic processing and decreased valuation of reward based on CAT (Lazarus, 1991; Smith and Ellsworth, 1985). It is found to insignificantly affect the delivery of online content (Berger and Milkman, 2012), and less likely to persuade readers to accept opinions. Accordingly, we assume the negative impact of disgust on review helpfulness.

Finally, joy and sadness elicited by the consumption of experience goods are possibly exaggerated, since extremely emotional reviews are perceived as fictitious (Banerjee et al., 2017; Xie et al., 2011). Sad individuals even tend to have trouble making rational decisions (Tiedens and Linton, 2001). Messages incorporating joy or sadness accordingly may lack comprehensive and detailed assessments of goods. We thus argue that joy and sadness have detrimental effects on review helpfulness. Hence, we propose the following hypotheses:

H1: Anger embedded in a review has a positive effect on perceived review helpfulness.
H2: Disgust embedded in a review has a negative effect on perceived review helpfulness.
H3: Fear embedded in a review has a positive effect on perceived review helpfulness.
H4: Joy embedded in a review has a negative effect on perceived review helpfulness.
H5: Sadness embedded in a review has a negative effect on perceived review helpfulness.
H6: Surprise embedded in a review has a positive effect on perceived review helpfulness.

Another important aim of this study is to examine the moderating effect of hotel star-classification on the relationships between emotions and review helpfulness. Anger and sadness in reviews tend to lead to extreme scores, and reviews with extreme ratings are considered more helpful than those with moderate ratings for low-quality goods and low-rated services (Shah and Lee, 2022; Yan et al., 2020). Extreme hotel reviews are more likely to be voted as helpful, working as reference points (Filieri et al., 2018). Thus, reviews incorporating anger or sadness are recognized as more helpful for low-rated hotels than high-rated hotel.

With regard to the rest of the emotions, we formulated hypotheses based on expectation-confirmation theory (Oliver, 1980). This theory proposes that consumers create expectations about product or service performance based on information they obtain. If the service performance is better than expected, consumers will be pleased and surprised. However, if the performance falls short of expectations, consumers will be discontented. Drawing upon the expectation-confirmation theory, we anticipate that the hotel star-rating moderates the impact of individual emotions on the review helpfulness. Specifically, as consumers have high expectations for high-rated hotels, they may not expect reviews with negative emotions (i.e., disgust and fear) and thus be more affected by them. On the contrary, consumers form low expectations of low-rated hotels and will be more affected by reviews incorporating positive emotions (i.e., joy and surprise). Therefore, we propose the following hypotheses:

H7: Hotel star-classification moderates the relationship between discrete emotions and review helpfulness in the following way.

H7a: Hotel star-classification negatively moderates the relationship between anger and review helpfulness.

H7b: Hotel star-classification positively moderates the relationship between disgust and review helpfulness.

H7c: Hotel star-classification positively moderates the relationship between fear and review helpfulness.

H7d: Hotel star-classification negatively moderates the relationship between joy and review helpfulness.

H7e: Hotel star-classification negatively moderates the relationship between sadness and review helpfulness.

H7f: Hotel star-classification negatively moderates the relationship between surprise and review helpfulness.

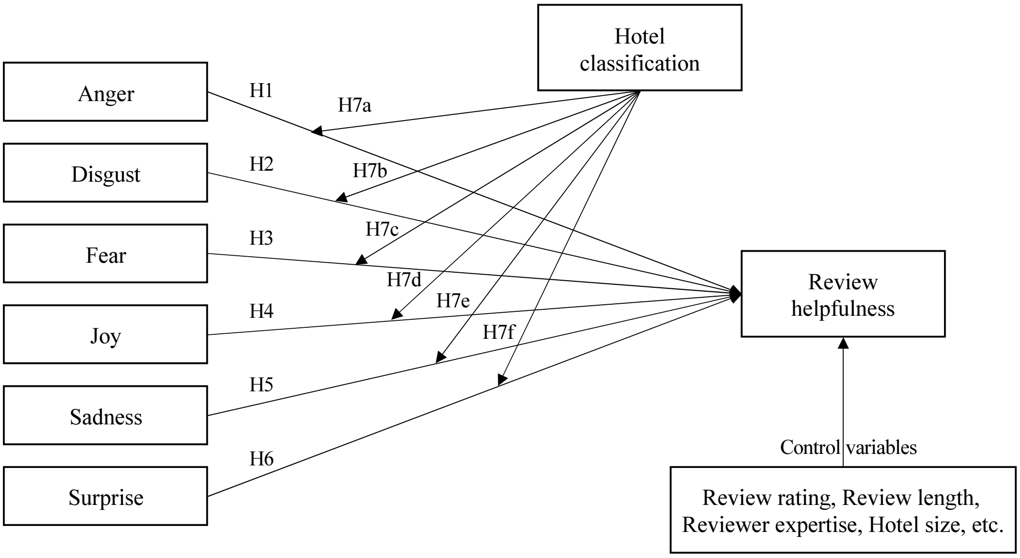
<Figure 1> presents the research framework proposed in this study.

Ⅳ. Methodology

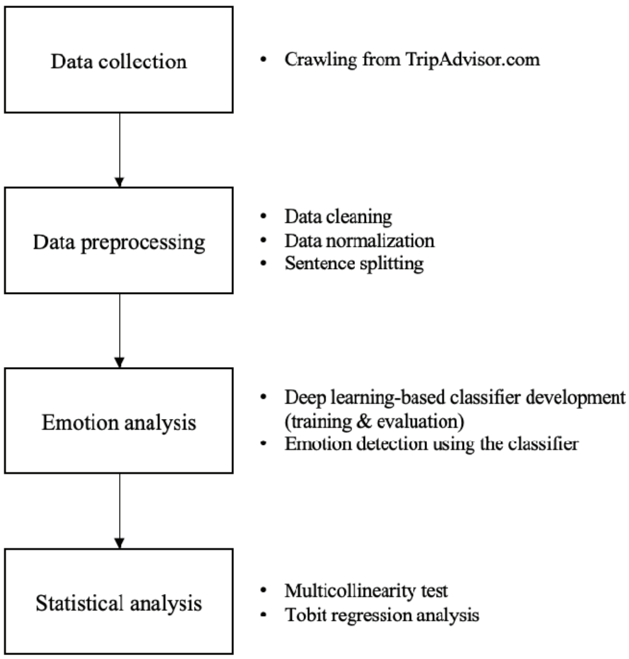
The research methodology consists of data collection followed by data preprocessing, emotion analysis, and statistical analysis, as depicted in <Figure 2>.

4.1. Data Collection and Preprocessing

First, we collected online reviews from TripAdvisor.com, concentrating on reviews of hotels located in New York City. New York City is considered a world mecca of entertainment and business and has been selected as the subject region for many previous studies (Lee et al., 2017; Rhee and Yang, 2015; Yang et al., 2021). Among 936 listed hotels, we randomly chose 200 and collected all reviews of these hotels written between January 2012 and December 2021 in English using a web crawler. We then excluded reviews with no helpful votes as in past research (Qazi et al., 2016; Ren and Hong, 2019) to reduce noise. Finally, a total of 56,157 reviews of 187 hotels was included in the data analysis. In the preprocessing stage, several steps were performed



<Figure 1> Research Framework



<Figure 2> Overview of Research Methodology

including data cleansing, transformation, and sentence splitting. We specifically adopted log transformation for the review length and the total number of reviews posted by a reviewer among all control variables to mitigate distributional skewness. We also applied the python library sentence splitter NLTK (Bird et al., 2009) to review texts including review title and content in order to analyze emotions per sentence.

4.2. Emotion Analysis

We constructed multi-class emotion classifiers using transfer learning methods to detect specific emotions in the review texts. Recently, language representation models to better understand the contextual meaning the words in text, such as Bidirectional Encoder Representations from Transformers (BERT; Devlin et al., 2018) and its variants, have been released and demonstrated state-of-the-art performances on various NLP tasks. We chose three models - RoBERTa (Liu et al., 2019), DistilBert (Sanh et al., 2019), and ELECTRA (Clark et al., 2020) - to train the classifiers.

To obtain a greater volume of training data for better classifier performance, we adopted a newly constructed dataset that is a combination of three different datasets for emotion detection, DailyDialog (Li et al., 2017), EmotionPush (Chen et al., 2018), and MELD (Poria et al., 2019). DailyDialog contains 102,979 utterances from 13,118 daily conversations, and EmotionPush includes 14,742 utterances from 1,000 Facebook messenger dialogues. MELD, which stands for Multimodal EmotionLines Dataset, consists of 13,708 utterances from 1,433 dialogues derived from the Friends TV series. These datasets were appropriate for this study because they cover every emotion class (a total of six classes) this research focuses on, as well as a neutral class to detect content with no emotional expressions. We specifically mixed datasets by each type of data – training, validation, and test, and only 20% of the neutral class data were selected by random sampling to be included in the final dataset so that they account for less than 50% percent of each type of dataset in order to resolve the problem of class imbalance. The statistics of the final dataset are described in <Table 1>, and the detailed statistics of the three source datasets (DailyDialog, EmotionPush, and MELD) are described in <Appendix A>. We found that the performance of the classifier can be improved by merging the datasets, as RoBERTa-based classifiers with the same hyperparameters (350, 4, 2e-5) showed an F1 of 60.35% solely with the MELD dataset, which has the least class imbalance among the three datasets,

<Table 1> Emotion Distribution in the Final Dataset

Emotion	Train	Validation	Test	Total
Anger	2,030	239	500	2,769
Disgust	659	31	130	820
Fear	450	55	69	574
Joy	14,407	1,007	1,879	17,293
Neutral	17,013	1,705	1,938	20,656
Sadness	2,041	228	397	2,666
Surprise	3,240	296	490	4,026
Total	39,840	3,561	5,403	48,804

but an F1 of 75.15% when trained and tested on the merged dataset, demonstrating an F1 difference of about 15%. The detailed performances of emotion classification on the MELD dataset are also presented in <Appendix B>.

We chose weighted average F1 and accuracy (micro-F1) scores as evaluation metrics (Hu et al., 2021; Kim and Vossen, 2021). During the fine-tuning process, we applied eights sets of hyperparameters (the maximum sequence length, the batch size, and the learning rate) based on the three transfer learning models and the number of epochs was set to 3. The performance results for the classifiers are presented in <Table 2>, which presents only the best performance of each model, with the entire evaluation results of all combinations of models and hyperparameters reported in <Appendix C>. The RoBERTa model with hyperparameters (350, 4, 2e-5) achieved the best performance with an F1-score of 75.15% and an accuracy of 75.26% and was selected to be implemented for emotion analysis.

After constructing the emotion classifier, we performed text mining to identify emotions embedded in the review texts. The classifier was applied to each sentence and determined whether any emotional state was expressed or not, and if so, which specific emotion was embedded in the corresponding sentence. <Table 3> shows some examples of emotions detected from each review sentence by the RoBERTa-based model and compares them to those detected by the NRC (Mohammad and Turney, 2010) dictionary-based model. When we performed emotion detection with the dictionary-based model, we excluded emotion categories that are not analyzed in this study (i.e., anticipation and trust) and employed NRCLex, a python package for classifying emotions based on the NRC lexicon (Mohammad and Turney, 2010), to sentences that are additionally preprocessed by lemmatization using NLTK (Bird et al., 2009). Our results indicate that unlike the dictionary-based model, the transfer learning model captures implicitly expressed emotions as it considers the contextual meaning brought by the negators (e.g., no, not) or the nuance of sentences, while the dictionary-based model does not. Through a supplementary experiment with the NRC lexicon-based method using the dataset used to generate transfer learning-based models, we also found that the lexicon-based method showed low accuracy (13.4%), again being unable to detect, for example, sad emotion in sentences with with negators such as "I'm not happy anymore."

4.3. Regression Analysis

4.3.1. Variables

The dependent variable in our conceptual model is review helpfulness, which is measured as the logarithmic form of the number of helpful votes received by online customer reviews because it has a skewed

<Table 2> Performance of the Emotion Classifiers

Model	Hyperparameters*	Accuracy	Precision	Recall	F1
RoBERTa	(350, 4, 2e-5)	75.26	75.61	75.56	75.15
DistilBERT	(350, 4, 2e-5)	73.86	73.30	73.88	73.29
ELECTRA	(350, 4, 2e-5)	74.98	74.55	74.88	74.85

Note: * Hyperparameters (x, y, z) for the fine-tuning, where x: max_seq_length (350, 500), y: train_batch_size (4, 8), and z: learning_rate (2e–5, 5e–5).

distribution <Table 4>.

The independent variables are the discrete six emotions embedded in the online customer reviews. They were measured according to the number of emotional sentences in the review, which were identified by transfer learning-based emotion analysis. Since the operational approach of measuring emotions as the proportion of emotional sentences in a review ((# emotional sentences / # sentences in a review) * 100) may cause multicollinearity when review length is included as a control variable (Ren and Hong, 2019), we counted the number of emotional sentences. To compare different distributions of emotions, the emotion variables are normalized via z-score.

Hotel star-rating was utilized as the moderator variable, measured by the number of stars assigned to each hotel and ranging from 1 to 5. The higher the rating, the better the quality of the hotel's customer service.

The control variables can be divided into review-related variables, reviewer-related variables, and hotel-related variables. Review rating, review rating squared, review length, and photos of the hotel represent the review features. Review rating is an overall rating that reviewers give to a hotel, and rating squared refers to a quadratic term of the rating value. Rating squared was additionally taken into account because of the nonlinear relationship between review ratings and helpfulness, following previous work (Filieri et al., 2018; Mudambi and Schuff, 2010; Ren and Hong, 2019; Yin et al., 2014). Review length was measured by word count. Photos of the hotel were operationalized by the number of photos posted with each online review. As mentioned in Section 4.1, we applied log transformation to the review length and the total number of reviews posted by a reviewer.

Concerning the reviewer-related variables, we included the total number of reviews posted by each reviewer and the total number of helpful votes ob-

<Table 3> Comparison of Emotions Detected by Two Different Models

Review Sentence	Emotion(s) Detected by Each Model	
	Transfer Learning Model (RoBERTa)	Dictionary-Based Model (NRC)
Every action this hotel takes is about stripping you of your dollar whilst providing next to no value for the pleasure. (Rating: 2)	Anger	Joy (*pleasure*)
Our room was dirty and so dusty. (Rating: 2)	Disgust	Disgust (dirty)
I was terrified until an employee told me that it does that often. (Rating: 3)	Fear	Neutral
Hyatt Place is almost always a cut above the competition and this location did not disappoint. (Rating: 5)	Joy	Anger (*disappoint*), disgust (*disappoint*), sadness (*disappoint*)
We wish we had booked somewhere else... (Rating: 1)	Sadness	Neutral
Above our expectations! (Rating: 4)	Surprise	Neutral

<Table 4> Operationalization of Variables

Variable Type	Variable	Operationalization	References
Dependent variable	Review helpfulness	A natural log transformation on the number of online users who voted a review as helpful	Filieri et al., 2021; Lee et al., 2017
Independent variables	Anger	A z-score normalization of (# anger-related sentences)	n.a.
	Disgust	A z-score normalization of (# disgust-related sentences)	n.a.
	Fear	A z-score normalization of (# fear-related sentences)	n.a.
	Joy	A z-score normalization of (# joy-related sentences)	n.a.
	Sadness	A z-score normalization of (# sadness-related sentences)	n.a.
	Surprise	A z-score normalization of (# surprise-related sentences)	n.a.
Moderator	Hotel star-rating (HSR)	The number of stars of a hotel	Filieri et al., 2021; Silva, 2015
Control variables	Review rating	The rating of the review expressed on a scale of 1-5	Filieri et al., 2018; Ren and Hong, 2019
	Review rating^2	The quadratic term of the star rating	Ren and Hong, 2019; Yang et al., 2020
	Review length	A natural log transformation on the number of words of a review	Liu and Park, 2015; Ren and Hong, 2019
	Photos of the hotel	The number of photos posted together with a review	Filieri et al., 2021; Lin et al., 2012
	Reviewer expertise	A natural log transformation on the number of reviews posted	Weiss et al., 2008
		The number of helpful votes obtained by the reviewer	Ghose and Ipeirotis, 2011
	Quality of the hotel	Certificate of excellence	Filieri et al., 2021; Kim et al., 2016
	Hotel size	The number of rooms in a hotel	Filieri et al., 2021; Simons and Hinkin, 2001

tained by the reviewer. Both variables capture the reviewer's expertise (Filieri et al., 2018).

To control for the effects of hotel-related features, we chose the certificate of excellence and hotel size as control variables. The certificate of excellence is measured using the presence or absence of a certificate of excellence assigned to each hotel by TripAdvisor, while the hotel size is measured by the number of rooms in each hotel.

4.3.2. Statistical Analysis

We employed Tobit regression for statistical testing following previous research (Filieri et al., 2021; Ren and Hong, 2019; Qazi et al., 2016). Tobit regression is appropriate for several reasons. First, the dependent variable (review helpfulness) has a censored nature, which means it has lower bounds. Second, the Tobit model can address potential sample selection problems, which means not all review readers have evaluated its helpfulness. We propose the two resulting equations, one of which is for testing hypotheses not concerning the moderating effect of the hotel star-rating (H1～H6) and the other for testing hypotheses addressing the moderating effect (H7).

$$ReviewHelpfulness_i = \beta_1 Anger_i + \beta_2 Disgust_i + \beta_3 Fear_i + \beta_4 Joy_i + \beta_5 Sadness_i + \beta_6 Surprise_i + \beta_7 ReviewRating_i + \beta_8 ReviewRating_i^2 + \beta_9 ReviewLength_i + \beta_{10} PhotosoftheHotel_i + \beta_{11} NumberofReviewsPosted_i + \beta_{12} NumberofHelpfulVotes\_ObtainedbytheReviewer_i + \beta_{13} QualityoftheHotel_i + \beta_{14} HotelSize_i + \beta_{15} HotelStarRating_i + \varepsilon_i \quad (1)$$

$$ReviewHelpfulness_i = \beta_1 Anger_i + \beta_2 Disgust_i + \beta_3 Fear_i + \beta_4 Joy_i + \beta_5 Sadness_i + \beta_6 Surprise_i + \beta_7 ReviewRating_i + \beta_8 ReviewRating_i^2 + \beta_9 ReviewLength_i + \beta_{10} PhotosoftheHotel_i + \beta_{11} NumberofReviewsPosted_i + \beta_{12} NumberofHelpfulVotes\_ObtainedbytheReviewer_i + \beta_{13} QualityoftheHotel_i + \beta_{14} HotelSize_i + \beta_{15} HotelStarRating_i + \beta_{16} Anger_i \times HSR_i + \beta_{17} Disgust_i \times HSR_i + \beta_{18} Fear_i \times HSR_i + \beta_{19} Joy_i \times HSR_i + \beta_{20} Sadness_i \times HSR_i + \beta_{21} Surprise_i \times HSR_i + \varepsilon_i \quad (2)$$

i refers to a given review and HSR is an abbreviation for hotel star-rating. In Equation (1), the impact of anger is estimated as β_1 to test the H1. In the same way, the impacts of the other emotions (disgust, fear, joy, sadness, and surprise) are estimated as each coefficient (β_2, β_3, β_4, β_5, and β_6) to test H2 to H6. Equation (2) was used to determine the coefficients of the variables -$Anger_i \times HSR_i$, $Disgust_i \times HSR_i$, $Fear_i \times HSR_i$, $Joy_i \times HSR_i$, $Sadness_i \times HSR_i$, and $Surprise_i \times HSR_i$) - to test H7a to H7f, respectively. We additionally tested Equation (1) with two different star-rating samples (low-rated group consisting of hotels with less than 4-star, and high-rated group of hotels with 4-star or above) to investigate changes in effect size of specific emotions on perceived review helpfulness.

Ⅴ. Results

The summary statistics for the variables are re-

<Table 5> Descriptive Statistics

Variable	Minimum	Maximum	Mean	Standard Deviation
Review helpfulness	0	5.642	0.301	0.442
Anger	-0.420	21.826	0	1
Disgust	-0.399	28.435	0	1
Fear	-0.145	36.379	0	1
Joy	-1.313	15.898	0	1
Sadness	-0.374	22.747	0	1
Surprise	-0.278	25.493	0	1
Review rating	1	5	4.247	1.159
Review rating^2	1	25	19.380	7.854
Review length	2.833	8.100	4.554	0.674
Photos of the hotel	0	3	0.232	0.728
Number of reviews posted	0	8.847	2.218	1.648
Helpful votes obtained by the reviewer	0	207,100	51.430	1281.112
Quality of the hotel	0	1	0.790	0.407
Hotel size	19	1,854	328.300	329.752
Hotel star-rating (HSR)	1	5	3.796	0.693

<Table 6> VIF and Tolerance Level Values

Variable	VIF	Tolerance
Anger	1.916	0.522
Disgust	1.635	0.612
Fear	1.084	0.922
Joy	1.760	0.568
Sadness	1.512	0.661
Surprise	1.146	0.873
Review rating	2.335	0.428
Review length	1.995	0.501
Photos of the hotel	1.056	0.947
Number of reviews posted	1.137	0.879
Helpful votes obtained by the reviewer	1.006	0.994
Quality of the hotel	1.231	0.812
Hotel size	1.114	0.897
Hotel star-rating (HSR)	1.171	0.854

ported in <Table 5>. Before testing the hypothetical relationships through regression analysis, a multicollinearity test was conducted to verify high correlations between the variables. The variance inflation factor (VIF) statistics of all independent variables were < 10 and the tolerance levels > 0.1, indicating multicollinearity was insignificant in our study <Table 6>.

The results of Tobit regression are presented in <Table 7>. There are two regression models, Model 1 for testing H1 to H6 and Model 2 for testing H7. The results indicate good fit, with highly significant likelihood ratios ($p < 0.001$) and pseudo R2 values of 0.195 and 0.197, respectively (Veall and Zimmermann, 1996). The control variables in both models showed significant impacts on review helpfulness. Review rating had negative effects on review helpfulness, while the remaining seven control variables positively influenced review helpfulness in this study.

Model 1 examines the direct effects of discrete emotions on review helpfulness. We found that four out of six emotions had significant impacts on perceived review helpfulness. Expressed anger, disgust, and joy had positive effects on review helpfulness, while fear negatively influenced review helpfulness. The influence of anger upon review helpfulness was stronger than those of disgust, fear, and joy, with $\beta = 0.022$ and p-value < 0.01. Therefore, H1, H2, H3, and H4 were supported. However, there were no significant effects of sadness or surprise on helpfulness votes, indicating that H5 and H6 were not supported.

In Model 2, we tested the moderating role of hotel star-rating on the relationship between specific emotions and review helpfulness. The hotel star-rating moderates the effects of three emotional dimensions (disgust, joy, and sadness) on perceived review helpfulness. It did not show significant moderating effects on the relationships between review helpfulness and any of the other three emotions including anger, fear, and surprise.

<Table 7> Tobit Regression Results for the Full Sample (N = 56,157)

Model	M1				M2			
Variable	**Coefficient**	**Std. error**	**t-value**	**Sig.**	**Coefficient**	**Std. error**	**t-value**	**Sig.**
Constant	-0.614	0.069	-8.888	0.000***	-0.612	0.069	-8.832	0.000***
Anger	0.022	0.007	3.341	0.001***	0.040	0.033	1.226	0.220
Disgust	-0.010	0.006	-1.688	0.092*	-0.125	0.031	-4.004	0.000***
Fear	0.011	0.005	2.244	0.025**	0.052	0.027	1.974	0.048**
Joy	-0.014	0.007	-2.029	0.042**	-0.079	0.030	-2.663	0.008***
Sadness	-0.005	0.006	-0.826	0.409	0.053	0.032	1.685	0.092*
Surprise	0.007	0.005	1.381	0.167	0.033	0.029	1.140	0.254
Review rating	-0.435	0.027	-16.374	0.000***	-0.436	0.027	-16.371	0.000***
Review rating^2	0.058	0.004	14.736	0.000***	0.058	0.004	14.758	0.000***
Review length	0.091	0.011	8.638	0.000***	0.090	0.011	8.549	0.000***
Photos of the hotel	0.037	0.007	5.410	0.000***	0.037	0.007	5.345	0.000***
Number of reviews posted	0.079	0.003	23.650	0.000***	0.079	0.003	23.631	0.000***
Helpful votes obtained by the reviewer	0.000	0.000	4.695	0.000***	0.000	0.000	4.707	0.000***
Quality of the hotel	0.217	0.014	15.518	0.000***	0.218	0.014	15.539	0.000***
Hotel size	0.000	0.000	15.857	0.000***	0.000	0.000	15.888	0.000***
Hotel star-rating (HSR)	0.048	0.008	6.217	0.000***	0.048	0.008	6.200	0.000***
Anger x HSR					-0.005	0.008	-0.560	0.575
Disgust x HSR					0.030	0.008	3.748	0.000***
Fear x HSR					-0.011	0.007	-1.572	0.116
Joy x HSR					-0.017	0.007	2.288	0.022**
Sadness x HSR					-0.015	0.008	-1.871	0.061*
Surprise x HSR					-0.007	0.007	-0.890	0.373
Log likelihood	-48758.000				-48747.060			
Pseudo R^2	0.195				0.197			

Note: *** p-value < 1%; ** p-value < 5%; * p-value < 10%.

To interpret the moderating effect of hotel-star rating more precisely, we further investigated the influence of discrete emotions by conducting regression analyses of two subsamples, low-rated hotels with less than 4 stars and high-rated hotels with 4 or more. According to <Table 8>, disgust embedded in the review had a greater negative effect on review helpfulness for low-rated hotels than high-rated hotels. Joy and sadness affected review helpfulness more negatively for high-rated hotels compared to low-rated hotels. Combined with the results confirmed by Model 2, the effect sizes of specific emotions including disgust, joy, and sadness on perceived review helpfulness were significantly moderated by ho-

<Table 8> Tobit Regression Results for Different Star-Rating Hotel Groups

Observations	Low-Rated Hotels (N = 20,415)				High-Rated Hotels (N = 35,742)			
Variable	**Coefficient**	**Std. error**	**T-Value**	**Sig.**	**Coefficient**	**Std. error**	**T-VALUE**	**Sig.**
Constant	-0.314	0.146	-2.149	0.032**	-1.548	0.100	-15.441	0.000***
Anger	0.026	0.011	2.371	0.018**	0.014	0.008	1.724	0.085*
Disgust	-0.039	0.010	-3.963	0.000***	0.005	0.008	0.705	0.481
Fear	0.013	0.007	1.827	0.068*	0.009	0.007	1.388	0.165
Joy	-0.003	0.011	-0.277	0.782	-0.016	0.008	-2.040	0.041**
Sadness	0.008	0.010	0.832	0.405	-0.016	0.007	-2.150	0.032**
Surprise	0.015	0.008	1.736	0.083*	0.004	0.006	0.690	0.490
Review rating	-0.436	0.042	-10.489	0.000***	-0.462	0.033	-13.845	0.000***
Review rating^2	0.059	0.006	9.709	0.000***	0.059	0.005	12.082	0.000***
Review length	0.081	0.017	4.726	0.000***	0.099	0.013	7.645	0.000***
Photos of the hotel	0.046	0.012	3.784	0.000***	0.032	0.008	3.997	0.000***
Number of reviews posted	0.083	0.006	15.000	0.000***	0.084	0.004	20.376	0.000***
Helpful votes obtained by the reviewer	0.000	0.000	2.952	0.003***	0.000	0.000	7.717	0.000***
Quality of the hotel	0.075	0.018	4.193	0.000***	0.165	0.022	7.356	0.000***
Hotel size	0.002	0.000	35.097	0.000***	0.000	0.000	3.026	0.002***
Hotel star-rating (HSR)	-0.106	0.035	-3.018	0.003***	0.298	0.015	20.271	0.000***
Log likelihood	-17277.240				-30727.270			
Pseudo R^2	0.452				0.285			

Note: *** p-value < 1%; ** p-value < 5%; * p-value < 10%.

tel star-rating, providing evidence to support H7b, H7d, and H7e. On the contrary, expressed anger, fear, and surprise seemed to have slightly greater positive impacts upon review helpfulness for low-rated hotels than high-rated hotels. H7a, H7c, and H7f were not supported because the hotel-star rating did not show any significant moderating effect on the relationship between those three emotional dimensions and review helpfulness as shown in Model 2.

Ⅵ. Discussion and Conclusion

6.1. Discussion of Findings

In this study, we investigated the differential effects of discrete emotions (anger, disgust, fear, joy, sadness, and surprise) on perceived review helpfulness. Through empirical research on hotel reviews, we determined that even though joy and surprise belong to the same valence, they have dissimilar impacts on perceived review helpfulness as stated in some prior studies (Shah and Lee, 2022; Wang et al., 2019). The findings of both this study and two prior studies (Shah and Lee, 2022; Wang et al., 2019) showed only joy has a significant negative impact on perceived review helpfulness, not surprise. However, Shah and Lee (2022) confirmed the negative effect of anger upon review helpfulness while other studies indicated positive effects.

We also validated the effect of the hotel star-classification as a moderator in the relationship between discrete emotions and review helpfulness. We found that the hotel star-rating moderates the effects of three emotions (disgust, joy, and sadness) on perceived review helpfulness. Specifically, we discovered joy had a greater negative impact on review helpfulness for high-rated hotels while disgust affected review helpfulness more negatively for low-rated hotels. This can be explained based on the expectation-confirmation theory (Oliver, 1980). Consumers formulate different expectations based on the information they obtain. The hotel class could be a credible information to form expectations of the hotel service performance. As hotel consumers are easily find a number of reviews incorporating positive emotions and form high expectations for high-rated hotels, the impact of joy would be attenuated. In contrast, consumers with low expectations for low-rated hotels are likely to be less influenced by disgust-embedded messages. Meanwhile, we found that sadness had a slightly greater positive impact for low-rated hotels. This may be due to the possible increase in helpful votes given to reviews with extreme ratings and expressed sadness, agreeing with the findings of prior research (Filieri et al., 2018; Shah and Lee, 2022; Yan et al., 2020). These findings are in line with the results of previous studies finding that the hotel category significantly influenced consumer decision making (Banerjee and Chua, 2019; Filieri et al., 2021; Galati and Galati, 2019).

6.2. Theoretical Implications

This research suggests a novel approach to measure emotion variables. By employing transfer learning-based emotion detection, it was possible to capture emotions implicitly expressed in the context of reviews, an innovation that complements the traditional dictionary-based approach. Previous studies (Ahmad and Laroche, 2015; Craciun et al., 2020; Yin et al., 2014) assumed that each emotion is expressed if a review includes corresponding emotional words based on open-source dictionaries the such as the NRC emotion lexicon (Mohammad and Turney, 2010) or LIWC dictionary (Pennebaker et al., 2015). Such dictionary-based methods cannot detect emotions when keywords co-occur with negators like ‘not’ or ‘no’. Additionally, they are largely dependent on the coverage of the dictionary used, so that emotion detection can vary depending on the recency of vocabulary coverage. However, we identified each emotion that the review writer intended to describe in review texts by applying a machine learning-based emotion classifier that specifically uses contextualized word embedding and was pre-trained using a large-scale text corpus.

To validate the effectiveness of emotions extracted by the classifier, we further conducted regression analysis on the associations of specific emotions and review ratings. As depicted in <Table 9>, we found that negative emotions such as anger and disgust had a tendency to associate with low rating grades, while positive emotions are related to high rating grades with the coefficient values which is positive or close to zero. The congruence between emotion valence and review rating demonstrates that the emotions recognized by the classifier are effective and implies that our suggested approach is appropriate to be applied in future research.

Another contribution is that our findings deepen the understanding of the effect of discrete emotions.

<Table 9> Tobit Regression Results with Review Rating as Dependent Variable

Dependent Variable	Review Rating			
Independent Variable	Coefficient	Std. error	T-Value	Sig.
Constant	5.350	0.011	468.170	0.000***
Anger	-0.567	0.010	-55.930	0.000***
Disgust	-0.552	0.009	-60.080	0.000***
Fear	-0.087	0.008	-10.470	0.000***
Joy	1.220	0.012	105.250	0.000***
Sadness	-0.301	0.009	-32.270	0.000***
Surprise	-0.051	0.009	-5.920	0.000***
Log likelihood	-54180.010			
Pseudo R^2	0.238			

Note: *** p-value < 1%.

Most previous literature on discrete emotions focused a few negative emotions (Craciun et al., 2020; Li et al., 2020; Ren and Hong, 2019; Yin et al., 2014) or analyzed different conditions or contexts of reviews such as product (Ahmad and Laroche, 2015; Craciun et al., 2020; Ren and Hong, 2019; Yin et al., 2014) or restaurant reviews (Li et al., 2020; Wang et al., 2019), resulting in the findings that were not conclusive. We used both positive and negative emotions as predictors of review helpfulness, and our results for hotel reviews generally support the findings of recent studies of the reviews of experience goods or services (Li et al., 2020; Wang et al., 2019), rather than those using the reviews of search goods (Ahmad and Laroche, 2015; Craciun et al., 2020; Ren and Hong, 2019; Yin et al., 2014). This emphasizes that it is critical to consider product type when analyzing the impacts of discrete emotions. Moreover, our research is the first to assess the moderating role of hotel classification in the relationship between emotions and review helpfulness. There has been little attention paid to the moderating effects of their relationship (Filieri et al., 2018; Karimi and Wang, 2017). This study highlights the importance of such effects in order to deeply understand the impact of emotional content.

6.3. Managerial Implications

Our study offers key insights for hotel managers who attempt to utilize online reviews to derive strategies for improving hotel service and reputation. Our empirical results will help managers to better understand the specific emotions that are influential on customers. Considering the finding that potential customers are especially concerned by anger and fear expressed in reviews, managers should pay careful attention to hotel attributes associated with those emotional dimensions. In particular, the managers of hotels with comparatively lower star-ratings should be extremely watchful regarding those features rather than attributes related to disgust, since for low-rated hotels, consumers care less about disgust expressed in reviews as it is expected to appear. On the other hand, managers of high-rated hotels should treat anger-related content more seriously compared to those with joy or sadness-related emotional words. We recommend that mangers to put more efforts into monitoring and responding to reviews in order to define affective features of the hotels, as most consumers are highly or fairly likely to use a business that responds to all of its online reviews (BrightLocal, 2022).

Second, this study also provides implications for the managers of review platforms about analyzing review data. For example, if they want to release new 'recommended' or 'most helpful' algorithms to sort reviews, they would be able to score reviews more elaborately by distinguishing reviews with the same number of helpful votes depending on embedded emotions. In addition, they could even highlight specific parts of the review, in which the emotional content is discovered, as helpful information for platform users. Such an approach would help review readers to more efficiently identify valuable reviews.

Finally, based on the findings of our study, the product mangers of e-commerce platforms can guide review writers to generate more helpful reviews. Since the length of reviews is usually limited, making the review content more helpful is essential to improve the perceived usefulness of review systems. For instance, managers can provide instructions for writing reviews for high-rated hotels that emphasize describing not only joyful experiences but anger-inducing experiences as well, since it was demonstrated that joy-embedded reviews have negative impacts on review helpfulness for hotels with high star-ratings.

6.4. Limitations and Future Research

There are some limitations of our research. Due to the nature of the dataset used in the study, the findings may not be generalized to data from other platforms like Amazon.com or regions such as European or Asian countries. The dataset of our study is based on reviews of hotels located in New York City, collected from TripAdvisor.com. In this vein, repeating the study with data across different platforms and regions might be promising. We also suggest that researchers examine whether our findings with our novel measurement method for emotion variables can be generalized to different types or services, as the hotel dataset represents only experience products, not search products. Furthermore, inclusion of other uninvestigated variables such as hotel price or review type can also extend the current research. Other factors that were not included as independent or control variables in this study may impact review helpfulness. Finally, the performance of our novel method to measure emotions based on machine learning should be further validated for improved performance. Even though our findings are in accordance with theory and with prior research, future research should apply the method to larger dataset or using data from other hotel review platforms to further validate our findings.

Acknowledgements

This research was conducted in 2020 with the support of the Ministry of Education and the Korea Research Foundation (NRF-2020S1A3A2A02093277).

<References>

[1] Ahmad, S. N., and Laroche, M. (2015). How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. *International Journal of Electronic*

Commerce, 20(1), 76-111. https://doi.org/10.1080/10864415.2016.1061471

[2] Baek, H., Ahn, J., and Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce, 17*(2), 99-126. https://doi.org/10.2753/JEC1086-4415170204

[3] Banerjee, S., and Chua, A. Y. (2019). Trust in online hotel reviews across review polarity and hotel category. *Computers in Human Behavior, 90*, 265-275. https://doi.org/10.1016/j.chb.2018.09.010

[4] Banerjee, S., Chua, A. Y., and Kim, J. J. (2017). Don't be deceived: Using linguistic analysis to learn how to discern online review authenticity. *Journal of the Association for Information Science and Technology, 68*(6), 1525-1538. https://doi.org/10.1002/asi.23784

[5] Berger, J., and Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192-205. https://doi.org/10.1509/jmr.10.0353

[6] Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with The Natural Language Toolkit.* O'Reilly Media, Inc.

[7] BrightLocal. (2022). Local consumer review survey 2022, Retrieved from https://www.brightlocal.com/research/local-consumer-review-survey/

[8] Callan, R. J. (1998). Attributional analysis of customers' hotel selection criteria by UK grading scheme categories. *Journal of Travel Research, 36*(3), 20-34. https://doi.org/10.1177/004728759803600303

[9] Cao, Q., Duan, W., and Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems, 50*(2), 511-521. https://doi.org/10.1016/j.dss.2010.11.009

[10] Chen, S. Y., Hsu, C. C., Kuo, C. C., and Ku, L. W. (2018). Emotionlines: An emotion corpus of multi-party conversations. *arXiv preprint. arXiv:1802.08379.*

[11] Chua, A. Y., and Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior, 54*, 547-554. https://doi.org/10.1016/j.chb.2015.08.057

[12] Clark, K., Luong, M. T., Le, Q. V., and Manning, C. D. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators. *arXiv preprint. arXiv:2003.10555.*

[13] Craciun, G., and Moore, K. (2019). Credibility of negative online product reviews: Reviewer gender, reputation and emotion effects. *Computers in Human Behavior, 97*, 104-115. https://doi.org/10.1016/j.chb.2019.03.010

[14] Craciun, G., Zhou, W., and Shan, Z. (2020). Discrete emotions effects on electronic word-of-mouth helpfulness: The moderating role of reviewer gender and contextual emotional tone. *Decision Support Systems, 130*, 113226. https://doi.org/10.1016/j.dss.2019.113226

[15] Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint. arXiv:1810.04805.*

[16] Dillard, J. P., Plotnick, C. A., Godbold, L. C., Freimuth, V. S., and Edgar, T. (1996). The multiple affective outcomes of AIDS PSAs: Fear appeals do more than scare people. *Communication Research, 23*(1), 44-72. https://doi.org/10.1177/009365096023001002

[17] Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6(3-4),* 169-200. https://doi.org/10.1080/02699939208411068

[18] Fan, W. (2021). What makes consumer perception of online review helpfulness: synthesizing the past to guide future research. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2738.

[19] Ferrer, R., Klein, W., Lerner, J., Reyna, V., and Keltner, D. (2016). Emotions and health decision making. In C. Roberto and I. Kawachi (Eds.), *Behavioral economics and Public Health* (pp. 101-132). Cambridge, MA: Harvard University Press.

[20] Filieri, R. (2015). What makes online reviews helpful?

A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of Business Research, 68*(6), 1261-1270. https://doi.org/10.1016/j.jbusres.2014.11.006

[21] Filieri, R., Galati, F., and Raguseo, E. (2021). The impact of service attributes and category on eWOM helpfulness: An investigation of extremely negative and positive ratings using latent semantic analytics and regression analysis. *Computers in Human Behavior*, *114*, 106527. https://doi.org/10.1016/j.chb.2020.106527

[22] Filieri, R., Raguseo, E., and Vitari, C. (2018). When are extreme ratings more helpful? Empirical evidence on the moderating effects of review characteristics and product type. *Computers in Human Behavior, 88*, 134-142. https://doi.org/10.1016/j.chb.2018.05.042

[23] Fontaine, J. R., Scherer, K. R., Roesch, E. B., and Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological science*, *18*(12), 1050-1057. https://doi.org/10.1111/j.1467-9280.2007.02024.x

[24] Galati, F., and Galati, R. (2019). Cross-country analysis of perception and emphasis of hotel attributes. *Tourism Management, 74*, 24-42. https://doi.org/10.1016/j.tourman.2019.02.011

[25] Garcia, D., and Schweitzer, F. (2011). Emotions in product reviews--empirics and models. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing* (pp. 483-488).

[26] Ghose, A., and Ipeirotis, P. G. (2010). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering, 23*(10), 1498-1512. https://doi.org/10.1109/TKDE.2010.188

[27] Hardeniya, T., and Borikar, D. A. (2016). Dictionary based approach to sentiment analysis-a review. *International Journal of Advanced Engineering, Management and Science*, *2*(5), 239438.

[28] Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., Liu, R., Ly, K., Kacker, V., She, B., and Li, Z. (2021). Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. *Journal of Medical Internet Research, 23*(9), e30854. https://doi.org/10.2196/30854

[29] Huang, A. H., Chen, K., Yen, D. C., and Tran, T. P. (2015). A study of factors that contribute to online review helpfulness. *Computers in Human Behavior, 48*, 17-27. https://doi.org/10.1016/j.chb.2015.01.010

[30] Janiesch, C., Zschech, P., and Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695.

[31] Jiang, Z., and Benbasat, I. (2007). The effects of presentation formats and task complexity on online consumers' product understanding. *Mis Quarterly*, 475-500.

[32] Karimi, S., and Wang, F. (2017). Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems, 96*, 39-48. https://doi.org/10.1016/j.dss.2017.02.001

[33] Khotimah, D. A. K., and Sarno, R. (2019). Sentiment analysis of hotel aspect using probabilistic latent semantic analysis, word embedding and LSTM. *International Journal of Intelligent Engineering and Systems*, *12*(4), 275-290.

[34] Kim, J., and Gupta, P. (2012). Emotional expressions in online user reviews: How they influence consumers' product evaluations. *Journal of Business Research*, *65*(7), 985-992. https://doi.org/10.1016/j.jbusres.2011.04.013

[35] Kim, T., and Vossen, P. (2021). EmoBERTa: Speaker-Aware Emotion Recognition in Conversation with RoBERTa. *arXiv preprint. arXiv:2108.12009.*

[36] Kim, W. G., Li, J. J., and Brymer, R. A. (2016). The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. *International Journal of Hospitality Management, 55*, 41-51. https://doi.org/10.1016/j.ijhm.2016.03.001

[37] Lazarus, R. S. (1991). *Emotion and Adaptation.*

Oxford University Press.
[38] Lee, M., Jeong, M., and Lee, J. (2017). Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: A text mining approach. *International Journal of Contemporary Hospitality Management, 29*(2), 762-783. https://doi.org/10.1108/IJCHM-10-2015-0626
[39] Lerner, J. S., and Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81*(1), 146. https://doi.org/10.1037/0022-3514.81.1.146
[40] Li, H., Liu, H., and Zhang, Z. (2020). Online persuasion of review emotional intensity: A text mining analysis of restaurant reviews. *International Journal of Hospitality Management, 89*, 102558. https://doi.org/10.1016/j.ijhm.2020.102558
[41] Li, Y., Su, H., Shen, X., Li, W., Cao, Z., and Niu, S. (2017). DailyDialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint. arXiv:1710.03957.*
[42] Lin, T. M., Lu, K. Y., and Wu, J. J. (2012). The effects of visual information in eWOM communication. *The Journal of Research in Indian Medicine, 6*(1), 7-26. https://doi.org/10.1108/17505931211241341
[43] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). RoBERTa: A robustly optimized bert pretraining approach. *arXiv preprint. arXiv:1907.11692.*
[44] Liu, Z., and Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management, 47*, 140-151. https://doi.org/10.1016/j.tourman.2014.09.020
[45] Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., and Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing, 77*(1), 87-103. https://doi.org/10.1509/jm.11.0560
[46] Ma, Y., Xiang, Z., Du, Q., and Fan, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning. *International Journal of Hospitality Management, 71*, 120-131. https://doi.org/10.1016/j.ijhm.2017.12.008
[47] Mohammad, S., and Turney, P. (2010). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text* (pp. 26-34).
[48] Mudambi, S. M., and Schuff, D. (2010). What makes a helpful review? A study of customer reviews on amazon. Com. *MIS Quarterly, 34*(1), 185-200.
[49] Nabi, R. (2002). Anger, fear, uncertainty, and attitudes: A test of the cognitive-functional model. *Communication Monographs, 69(3),* 204-216. https://doi.org/10.1080/03637750216541
[50] Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research, 17*(4), 460-469. https://doi.org/10.1177/002224378001700405
[51] Otterbacher, J. (2009). 'Helpfulness' in online communities: a measure of message quality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 955-964).
[52] Pan, Y., and Zhang, J. Q. (2011). Born unequal: A study of the helpfulness of user-generated product reviews. *Journal of Retailing, 87*(4), 598-612. https://doi.org/10.1016/j.jretai.2011.05.002
[53] Park, D. H., and Lee, J. (2008). EWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications, 7*(4), 386-398. https://doi.org/10.1016/j.elerap.2007.11.004
[54] Peng, C., Yin, D., Wei, C., and Zhang, H. (2014). How and when review length and emotional intensity influence review helpfulness: Empirical evidence from epinions.com. In *Proceedings of Thirty Fifth International Conference on Information Systems* (pp. 1-16).

[55] Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). *Linguistic Inquiry and WORD COUNT: LIWC2015*. Austin, TX: Pennebaker Conglomerates.

[56] Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., and Mihalcea, R. (2018). MELD: A multimodal multi-party dataset for emotion recognition in conversations. arXiv preprint. arXiv: 1810.02508.

[57] Provost, F., and Kohavi, R. (1998). Glossary of terms. *Journal of Machine Learning*, *30*(2-3), 271-274. https://doi.org/10.1023/A:1017181826899

[58] Qazi, A., Syed, K. B. S., Raj, R. G., Cambria, E., Tahir, M., and Alghazzawi, D. (2016). A concept-level approach to the analysis of online review helpfulness. *Computers in Human Behavior, 58*, 75-81. https://doi.org/10.1016/j.chb.2015.12.028

[59] Racherla, P., and Friske, W. (2012). Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications, 11*(6), 548-559. https://doi.org/10.1016/j.elerap.2012.06.003

[60] Ren, G., and Hong, T. (2019). Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information Processing and Management*, *56*(4), 1425-1438. https://doi.org/10.1016/j.ipm.2018.04.003

[61] Rhee, H. T., and Yang, S. B. (2015). Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings. *Computers in Human Behavior*, *50*, 576-587. https://doi.org/10.1016/j.chb.2015.02.069

[62] Rozin, P., and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296-320. https://doi.org/10.1207/S15327957PSPR0504_2

[63] Salehan, M., and Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, 30-40. https://doi.org/10.1016/j.dss.2015.10.006

[64] Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint. arXiv: 1910.01108*.

[65] Shah, A. M., and Lee, K. (2022). The role of emotions intensity in helpfulness of online physician reviews. *Intelligent Automation and Soft Computing, 31(3)*, 1719-1735. https://doi.org/10.32604/iasc.2022.019666

[66] Silva, R. (2015). Multimarket contact, differentiation, and prices of chain hotels. *Tourism Management, 48*, 305-315. https://doi.org/10.1016/j.tourman.2014.11.006

[67] Simons, T., and Hinkin, T. (2001). The effect of employee turnover on hotel profits: A test across multiple hotels. *Cornell Hotel and Restaurant Administration Quarterly, 42(4)*, 65-69. https://doi.org/10.1016/S0010-8804(01)80046-X

[68] Smith, C. A., and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of ,ersonality and Social Psychology*, *48*(4), 813-838. https://doi.org/10.1037/0022-3514.48.4.813

[69] Sniezek, J. A., and Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, *84*(2), 288-307. https://doi.org/10.1006/obhd.2000.2926

[70] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., and Liu, C. (2018). A survey on deep transfer learning. *International Conference on Artificial Neural Networks*, 270-279.

[71] Tiedens, L. Z., and Linton, S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social Psychology*, *81*(6), 973-988. https://doi.org/10.1037/0022-3514.81.6.973

[72] Veall, M. R., and Zimmermann, K. F. (1996). Pseudo-R2 measures for some common limited dependent variable models. *Journal of Economic Surveys, 10(3)*, 241-259. https://doi.org/10.1111/j.1467-6419.1996.t

b00013.x

[73] Wang, X., Tang, L. R., and Kim, E. (2019). More than words: Do emotional content and linguistic style matching matter on restaurant review helpfulness? *International Journal of Hospitality Management*, *77*, 438-447. https://doi.org/10.1016/j.ijhm.2018.08.007

[74] Weiss, A. M., Lurie, N. H., and MacInnis, D. J. (2008). Listening to strangers: whose responses are valuable, how valuable are they, and why? *Journal of marketing Research*, *45*(4), 425-436. https://doi.org/10.1509/jmkr.45.4.425

[75] Westbrook, R. A., and Oliver, R. L. (1991). The dimensionality of consumption emotion patterns and consumer satisfaction. *Journal of Consumer Research*, *18*(1), 84-91. https://doi.org/10.1086/209243

[76] Xie, H. J., Miao, L., Kuo, P. J., and Lee, B. Y. (2011). Consumers' responses to ambivalent online hotel reviews: The role of perceived source credibility and pre-decisional disposition. *International Journal of Hospitality Management*, *30*(1), 178-183. https://doi.org/10.1016/j.ijhm.2010.04.008

[77] Yan, X., Khan, S., and Shah, S. J. (2020). Exploring the impact of review and service-related signals on online physician review helpfulness: A multi-methods approach. In *Proceedings of Twenty-Fourth Pacific Asia Conference on Information Systems* (pp. 1-14).

[78] Yang, S., Zhou, C., and Chen, Y. (2021). Do topic consistency and linguistic style similarity affect online review helpfulness? An elaboration likelihood model perspective. *Information Processing and Management, 58*(3), 102521. https://doi.org/10.1016/j.ipm.2021.102521

[79] Yang, S., Yao, J., and Qazi, A. (2020). Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining. *Information Processing and Management, 57*(2), 102179. https://doi.org/10.1016/j.ipm.2019.102179

[80] Yin, D., Bond, S. D., and Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*(2), 539-560.

[81] Yin, D., Bond, S. D., and Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research*, *54*(3), 447-463. https://doi.org/10.1509/jmr.13.0379

[82] Zhou, S., and Guo, B. (2017). The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93, 77-87. https://doi.org/10.1016/j.dss.2016.09.016

[83] Zhu, L., Yin, G., and He, W. (2014). Is this opinion leader's review useful? Peripheral cues for online review helpfulness. *Journal of Electronic Commerce Research*, *15*(4), 267-280.

<Appendix A> Statistics of the Source Datasets for Training and Testing the Transfer Learning Models

Emotion	DailyDailog			EmotionPush			MELD			Total
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test	
Anger	827	77	118	94	9	37	1,109	153	345	2,769
Disgust	303	3	47	85	6	15	271	22	68	820
Fear	146	11	17	36	4	2	268	40	50	574
Joy	11,182	684	1,019	1,482	160	458	1,743	163	402	17,293
Neutral	72,143	7,108	6,321	8,212	946	2,115	4,710	470	1,256	20,656*
Sadness	969	79	102	389	38	87	683	111	208	2,666
Surprise	1,600	107	116	435	39	93	1,205	150	281	4,026
Total	87,170	8,069	7,740	10,733	1,202	2,807	9,989	1,109	2,610	48,804

*Note: * Only 20% of the merged data were included by random sampling to resolve the problem of class imbalance.*

<Appendix B> Performance of the Emotion Classifiers on the MELD Dataset

Model	Hyperparameters*	Accuracy	Precision	Recall	F1
RoBERTa	(350, 4, 2e-5)	62.13	**60.12**	62.23	**60.35**
	(350, 4, 5e-5)	60.41	57.82	60.38	58.35
	(350, 8, 2e-5)	**62.22**	59.47	**62.38**	59.85
	(350, 8, 5e-5)	60.69	58.30	60.79	58.88
	(500, 4, 2e-5)	58.61	57.53	58.68	57.83
	(500, 4, 5e-5)	59.33	57.04	59.20	57.47
	(500, 8, 2e-5)	61.77	59.62	61.62	59.96
	(500, 8, 5e-5)	61.14	58.44	61.18	59.09
DistilBERT	(350, 4, 2e-5)	59.87	57.01	59.99	57.33
	(350, 4, 5e-5)	57.80	56.12	57.74	56.30
	(350, 8, 2e-5)	60.05	57.74	60.16	57.50
	(350, 8, 5e-5)	59.96	58.18	59.96	57.71
	(500, 4, 2e-5)	60.23	56.69	60.05	57.42
	(500, 4, 5e-5)	57.80	55.88	57.70	55.86
	(500, 8, 2e-5)	59.87	58.54	59.92	57.41
	(500, 8, 5e-5)	60.14	57.35	60.14	57.76
ELECTRA	(350, 4, 2e-5)	61.95	58.88	61.86	59.00
	(350, 4, 5e-5)	60.14	56.55	60.37	57.81
	(350, 8, 2e-5)	61.59	57.33	61.75	59.14
	(350, 8, 5e-5)	42.38	17.80	42.38	25.43
	(500, 4, 2e-5)	60.32	59.60	60.50	57.83
	(500, 4, 5e-5)	59.96	58.43	59.79	58.79
	(500, 8, 2e-5)	61.14	56.75	61.09	58.50
	(500, 8, 5e-5)	42.38	17.80	42.38	25.43

*Note: * Hyperparameters (x, y, z) for the fine-tuning, where x: max_seq_length (350, 500), y: train_batch_size (4, 8), and z: learning_rate (2e–5, 5e–5).*

<Appendix C> Performance of the Emotion Classifiers on the Final Dataset

Model	Hyperparameters*	Accuracy	Precision	Recall	F1
RoBERTa	(350, 4, 2e-5)	**75.26**	**75.61**	**75.56**	**75.15**
	(350, 4, 5e-5)	74.28	74.01	74.41	73.85
	(350, 8, 2e-5)	74.84	75.11	74.71	74.72
	(350, 8, 5e-5)	74.14	74.34	74.27	73.99
	(500, 4, 2e-5)	74.30	74.48	74.32	74.07
	(500, 4, 5e-5)	73.55	73.77	73.53	73.42
	(500, 8, 2e-5)	74.30	74.48	74.32	74.07
	(500, 8, 5e-5)	73.83	73.95	73.95	73.70
DistilBERT	(350, 4, 2e-5)	73.86	73.30	73.88	73.29
	(350, 4, 5e-5)	72.73	72.67	72.65	72.09
	(350, 8, 2e-5)	73.10	72.85	72.75	72.38
	(350, 8, 5e-5)	72.65	72.04	72.59	72.11
	(500, 4, 2e-5)	73.49	72.93	73.69	73.12
	(500, 4, 5e-5)	72.23	71.94	72.10	71.48
	(500, 8, 2e-5)	73.52	72.86	73.32	73.07
	(500, 8, 5e-5)	72.56	72.32	72.49	71.96
ELECTRA	(350, 4, 2e-5)	74.98	74.55	74.88	74.85
	(350, 4, 5e-5)	74.19	73.97	74.38	73.83
	(350, 8, 2e-5)	74.64	74.24	74.76	73.88
	(350, 8, 5e-5)	74.25	74.34	74.38	73.81
	(500, 4, 2e-5)	74.11	74.05	73.98	73.74
	(500, 4, 5e-5)	47.88	22.98	47.88	31.12
	(500, 8, 2e-5)	73.83	73.87	73.77	73.64
	(500, 8, 5e-5)	73.24	72.37	73.23	72.48

*Note: * Hyperparameters (x, y, z) for the fine-tuning, where x: max_seq_length (350, 500), y: train_batch_size (4, 8), and z: learning_rate (2e-5, 5e-5).*

◆ About the Authors ◆

Dahee Lee

Dahee Lee is working at the IT product planning division of LG Uplus. She is a Ph.D. candidate in the School of Business at Hanyang University in Korea. She received a B.A. degree both from the Department of Library and Information Science and the Department of Business Administration at Yonsei University in Korea. She has a M.A. degree in Library and Information Science from Yonsei University in Korea. Her research interests include text mining and E-commerce. Her papers have been published in *PloS One*, *Scientometrics*, *Journal of Biomedical Informatics*, *Journal of Alzheimer's Disease*, and *Journal of Information Science*.

Jong Woo Kim

Jong Woo Kim is a professor at the School of Business, Hanyang University, Seoul, Korea. He received B.S. degree from the Department of Mathematics at Seoul National University, Seoul, Korea. He received his M.S. and Ph.D. degrees, respectively, from the Department of Management Science, the Department of Industrial Management at Korea Institute of Science and Technology (KAIST), Korea. His current research interests include intelligent information systems, data mining applications, text mining application, sentiment and emotion analysis, social network analysis, collaborative systems, and e-commerce recommendation systems. His papers have been published in *Expert Systems with Applications*, *Cyberpsychology Behavior and Social Networking*, *Computers in Human Behavior*, *Information Systems Frontiers*, *International Journal of Electronic Commerce*, *Electronic Commerce Research*, *Mathematical and Computer Modeling*, *Journal of Intelligent Information Systems*, and other journals.

Submitted: October 25, 2023; 1st Revision: December 27, 2023; Accepted: January 18, 2024