IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 24, 2022, accepted April 10, 2022, date of publication April 14, 2022, date of current version April 26, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3167511

Do Reviewers' Words and Behaviors Help Detect Fake Online Reviews and Spammers? Evidence From a Hierarchical Model

THI-KIM-HIEN LE[1,2,3], YI-ZHEN LI[1], AND SHENG-TUN LI[1,4,5]

[1]Institute of Information Management, National Cheng Kung University, Tainan 70101, Taiwan
[2]University of Economics and Law, Ho Chi Minh city 700000, Vietnam
[3]Vietnam National University, Ho Chi Minh city 700000, Vietnam
[4]Center for Innovative FinTech Business Models, National Cheng Kung University, Tainan 70101, Taiwan
[5]Department of Industrial and Information Management, National Cheng Kung University, Tainan 70101, Taiwan

Corresponding author: Sheng-Tun Li (stli@mail.ncku.edu.tw)

This work was supported in part by the Ministry of Science and Technology, Taiwan, under contract MOST 107-2410-H-006-046 and MOST 108-2410-H-006-106-MY3 and in part by the University of Economics and Law, Vietnam National University Ho Chi Minh city.

ABSTRACT Although numerous studies have investigated spam detection and spammer detection on online platforms, they have ignored the fact that reviews written by the same reviewer may be correlated because each reviewer has their own distinct style. The traditional logistic regression model cannot handle this type of data because they violate the independence of residuals assumption. Furthermore, relatively few studies related to fake review detection have considered linguistic and behavioral aspects simultaneously. Thus, we propose a hierarchical logistic regression (HLR)-based model for detecting fake reviews that considers both linguistic and behavioral characteristics. With this outcome, our kernel also has multiple applications, including the detection of review spammers as a pre-module of quality in machine learning. The experimental results demonstrate that HLR can classify fake reviews and review spammers more accurately than the standard machine-learning algorithms.

INDEX TERMS Fake reviews detection, review spammers detection, hierarchical logistic regression model, reviewer behaviors, linguistic styles.

I. INTRODUCTION

The rapid development of the Internet has led to the rising availability of review services on online platforms, such as shopping websites (e.g., amazon.com) and opinion-sharing websites (e.g., epinions.com). People typically do not now purchase products or services without first reading the reviews [1]; consumer-generated reviews have become an indispensable part of the online shopping experience. However, approximately 30% of online reviews are fake [2] and can mislead consumers into making poor decisions. They may even undermine the credibility and usefulness of reviews in general. Positive fake reviews can boost sales, conferring prestige and financial benefits on both the corporate and individual levels. By contrast, negative fake reviews can exert severe negative effects on the sales of a product or service and may even threaten the reputation of the relevant firm [3]. Therefore, the detection and elimination of review spam is essential for protecting the interests of consumers and sellers alike.

Unlike other kinds of spam (e.g., Web spam or email spam), review spam (i.e., fake reviews) is considerably more challenging to detect. Specifically, human users experience difficulty recognizing review spammers because spammers can easily pretend to be legitimate reviewers. Furthermore, given the openness of product review sites, spammers can pose as numerous users, thereby complicating their eradication. Some paid professionals fabricate reviews without having used the product or service in question. Their sole goal is to promote the reputation of their employer or undermine that of their employer's competitors [4], [5]. Such behaviors undercut the credibility of review platforms. In sum, distinguishing fake online reviewers from real ones is a formidable challenge because review spammers can outsmart genuine users by mimicking their behavior [6].

Fake review detection and review spammer detection have been investigated for many years. [7] conducted the first study on spam detection in which they constructed a classifier that

The associate editor coordinating the review of this manuscript and approving it for publication was Fu Lee Wang.

employed certain types of duplicate reviews as positive training data; the remaining types were used as negative training data. A more in-depth scrutiny was continued performing by them in [8], this study contributed to the identification of three types of spam reviews, namely untruthful opinions, reviews on brands only, and non-reviews by representing a review using a combination of review, reviewer and product-level features. [9] developed a scoring method for measuring the degree of spam produced by each reviewer by examining several behavioral characteristics of spammers. Subsequently, they modeled these behaviors for spammer detection. To detect deceptive opinions, [10] explored psycholinguistic features by combining the Linguistic Inquiry and Word Count (LIWC) text analysis program with analysis of standard words and part of speech (POS) n-gram features. Many other studies on spam detection and spammer detection have been done in recent years, for example [11]–[14].

Although numerous studies on identifying fake reviews and review spammers within online platforms have been performed, the nested relationship between reviews and reviewers has not been investigated. Existing standard single-level models treat all electronic word of mouth reviews as independent observations. For example, [15] employed logistic regression to determine whether reviews were manipulative or authentic by considering linguistic cues, such as the readability, genre, and writing style of negative reviews. [7] conducted logistic regression by using a data set obtained by crawling amazon.com. Review content and reviewer-specific features were extracted with 78% accuracy. Notably, reviews written by the same reviewer may be correlated with each other because each reviewer expresses their knowledge and ideas through their own distinct style [16], [17]. This correlation severely contravenes one of the most pivotal assumptions of conventional regression analysis, namely the assumption of independence of the residuals [18]. This violation can lead to underestimation of the standard error, which in turn can contribute to incorrect findings of significance. Furthermore, variability among reviewers and among reviews nested within reviewer clusters cannot be resolved by single-level logistic models.

Another shortcoming of the literature on fake review and review spammer detection is that linguistic features (e.g., LIWC, POS tagging) and behavioral features (e.g., life tenure, rating deviation) have rarely been examined simultaneously. For example, [19] incorporated linguistic features of reviews involving POS tagging, unigram, and LIWC analyses into their forecasting model and reported a detection accuracy of 65%. [20] examined reviewer features (e.g., the average proportion of unhelpful reviews and the ratio of the number of first reviews of a product to the total number of reviews written by that reviewer), achieving a precision of 64%. These findings provide compelling evidence to support the premise that both linguistic and behavioral characteristics can play integral roles in the detection of fake reviews and review spammers. However, few studies have probed both types of features in an integrated manner, despite the fact that combining them typically yields more favorable detection performance than considering each type of characteristic separately. For example, by using integrated features from review and reviewer, [21] attained accuracy of 90% in detecting fake reviewers, [1] achieved an F1-score of 95% in spammers detection, and [22] identified fake reviews with 95% accuracy. Therefore, there is a need to construct a fake-review-predicting model, with features from both categories being used simultaneously.

To detect fake reviews, we developed a hierarchical logistic regression (HLR) model that takes into account the characteristics of both reviews and reviewers. Hierarchical models carefully consider variability at each level of the hierarchy. Specifically, they enable cluster effects at different levels to be analyzed by providing estimates of how much of the variance is attributable to the reviewer and to each individual review. Hierarchical modeling is a highly recommended statistical method for handling this type of data structure when individual reviews are grouped within reviewer clusters. Because machine-learning algorithms cannot handle nesting, our kernel has various applications, including the detection of review spammers as a premodule of quality in machine learning. For model validation, we conducted an experiment into two parts: recency analysis and duration analysis. In both parts, we first conducted HLR under the consideration of both linguistic and behavioral features. We then used the results as inputs for detecting review spammers. The experimental results demonstrated that the HLR model was more effective in differentiating between fake and real reviews and reviewers than were the logistic regression (LR) algorithm and other machine-learning algorithms, namely support vector machine (SVM), random forest (RF), naive Bayes (NB), and k-nearest neighbor (KNN). The highest accuracy rate of fake review detection and review spammer detection was 86% and 94%, respectively. The remainder of this paper is organized as follows. Section II provides the theoretical foundation. Section III introduces the methods and research process. Sections IV and V detail the experiments and results. In Section VI, the conclusions and future directions are presented.

II. THEORETICAL FOUNDATION

In this section, we present the theoretical foundation of fake review detection employed in this study, namely linguistic analysis, the behavioral features of reviewers, and HLR.

A. LINGUISTIC ANALYSIS

Studies have established that writing is a stable, reliable, and personalized trait. Text analysis programs can be used to link natural language characteristics to personality characteristics [10], [17], [23]. Moreover, individuals express their knowledge and ideas through a unique linguistic style [16], [24]. In the analysis of fake reviews, the most widely used linguistic analysis features include LIWC, readability, and POS. Herein, we further applied two features that had not

previously been employed in relevant research—evidentiality and credibility.

1) LIWC FACTORS

LIWC is a text analysis tool that links natural word use to personality traits [25]. Through a psychological approach, it counts words within a given text sample irrespective of the context in which the words occur. The LIWC dictionary contains 80 psychological categories into which 4500 words are classified. LIWC enables researchers to explore the linguistic features of textual data, such as the number of pronouns and numbers of positive and negative emotion words. Using the LIWC engine, users can create their own internal dictionary with which to analyze text files and dimensions of interest [26]. [10] achieved more favorable results when they considered LIWC attributes alongside the bag-of-words model than with the bag-of-words model alone. In the context of online communities in which opinions and experiences are shared, the LIWC tool facilitates analysis of reviewers' personal traits and fraudulent behavior [10], [23].

2) READABILITY

Although the literature offers various definitions, readability generally refers to the quality that makes some texts easier to read than others [27]. One study stated that readability is the speed at which a text can be read as well as the ease with which the content can be understood and retained [28]. Focusing on the issue of writing style, [29] defined readability as ''the ease of understanding or comprehension due to the style of writing,'' not relating readability to the content, coherence, or organization of a text. In the present context, readability represents the effort and expertise a person requires to comprehend a review [30], [31]. [15] indicated that fake reviews have higher readability than do genuine reviews because review spammers use words that are easier to understand such that the review can be read rapidly. In line with relevant studies, we employed readability as a linguistic feature through which fake reviews and review spammers were identified.

3) CREDIBILITY

Credibility refers to the quality and professionalism of a review [32]. Various studies have considered credibility in text analysis. For example, [33] took credibility into account to compile a list of indicators for assessing the credibility of blogs. The indicators comprised several main categories: information quality, appeals and triggers of a personal nature, the blogger's expertise and disclosure of their offline identity, and the blogger's profile and value system. [34] asserted that customers tend to vote for reviews that direct them to credible sources; as a result, their helpfulness ratings are increased. The work of [35] also used credibility to discover the major elements attributing to review helpfulness, among which the average user helpfulness, the number of user reviews, the average business helpfulness, and the review length were of the utmost significance. [17] used credibility to predict the usefulness of reviews. The indicators examined were correct capitalization, emoticons, all capitals (which suggests a ''shouting'' tone), and misspelling. The researchers determined that credible reviews employ correct capitalization (including by minimizing the use of all capitals) and contain fewer emoticons and misspelled words. Given its demonstrated ability to predict review helpfulness, we believe that credibility also plays a vital role in identifying fake reviews and review spammers. Therefore, we incorporated it as a linguistic feature in our detection model.

4) EVIDENTIALITY

[36] defined evidentiality as the linguistic representation of evidence for a statement and its use as an explicit linguistic system to indicate the quality of information; evidentiality offers evidence to aid direct determination of the trustworthiness of a text [37], [38]. The linguistic definition of evidentiality has two dimensions: 1) as a label to indicate the source of information on narrated events [38] and 2) the evidence with which information is obtained [39]. [37] devised a linguistic model in which the concept of evidentiality was incorporated into a machine learning–based text classification framework. Evidential information provides clues instrumental to predicting the value of a text. Because the objective of review spammers is to convince other users to agree with their opinions, their reviews tend to be clearer and more straightforward than genuine reviews. Accordingly, evidentiality should be adaptable to the evaluation of fake reviews and review spammers.

5) POS

POS tagging, which provides syntactic (or grammatical) information on a sentence, has been used in natural language processing to measure text informativeness [40]. This method relies on the assumption that spammers cannot replicate all aspects of natural language when repurposing content [41]. A study by [10] reported that fake reviews contain more verbs, adjectives, and superlatives than do genuine reviews because review spammers tend to exaggerate. [42] employed supervised algorithms in classifying a data set of fake reviews. The researchers also performed POS tagging, focusing on the writing style. They detected fake reviews with outstanding accuracy (91.51%). Following these studies, we applied POS as linguistic features for identifying fake reviews and review spammers.

B. BEHAVIORAL FEATURES

1) RATING ENTROPY

Introduced by Claude Shannon in 1948, the concept of entropy holds that the average level of uncertainty inherent in a variable's probable outcomes is the entropy of a random variable [43]. On this basis, we defined rating entropy as the average level of disorder in a reviewer's rating scores. Genuine reviewers are more likely to base their reviews on merit, resulting in balanced reviews—that is, reviews that are

equally critical and noncritical. Spammers, by contrast, are likely to present extreme opinions given that their objective is either to artificially increase the ranking of a product or service or to lower the ranking of its competitors. According to [2], suspicious reviews are more extreme than genuine evaluations. [44] noted that review spammers tend to leave extreme ratings. Thus, we considered entropy a behavioral feature useful for recognizing fake reviews and review spammers.

2) RATING DEVIATION

Rating deviation refers to the amount by which a reviewer's rating of a product or service differs from the average rating of that product or service. In general, genuine reviewers rate a product comparably to other reviewers. By contrast, spammers are more likely to give low-quality products high ratings and give high-quality products low ratings to promote and undermine these products, respectively. [9] observed that spammers tend to deviate from the general rating consensus. The more a reviewer's rating of a product differs from the average rating, the greater that reviewer's rating deviation. [45] indicated that rating deviation is among the most critical features for identifying fake reviews. Notably, rating deviation provides clues pivotal to determining the quality of a review or reviewer.

3) REVIEW COUNT

Review count is the number of reviews written by a particular reviewer. This is a valuable factor and helps distinguish between review spammers and genuine reviewers. Specifically, spammers can make more reviews than genuine reviewers. In some circumstances, spammers evade detection or blacklisting by posting a small number of reviews from one account and then creating a new account from which to post more reviews. Studies on review manipulation have demonstrated that the behavioral distributions of opinion spammers differ from those of non-spammers and that publishing numerous reviews signals deviant behavior [44], [46]. Therefore, we considered review count a useful behavioral feature for recognizing fake reviews and review spammers.

4) LIFE TENURE

Life tenure is defined as the duration for which a reviewer has been active in an online forum [47], [48]. Specifically, studies have indicated that users' confidence regarding the authenticity of a reviewer increases with the amount of time that reviewer has been active in the forum or on the review website [49], [50]. According to [22], real consumers use their accounts to post reviews occasionally, whereas review spammers remain members of a platform for only a short period and post a relatively high number of reviews in that period. In sum, one individual activating in a short time and post numerous reviews is indicative of suspicious behavior. By contrast, a reviewer who is visible for a relatively long period and post review periodically corresponds to normal behavior [46]. This is why we used life tenure as a feature for recognizing review spammers.

5) REVIEW GAP

The average time between one reviewer's successive comments is known as the review gap. This is a useful metric for identifying potential review spammers who are likely to try to copy the average person's review-posting frequency. Research has demonstrated that some people send messages in bursts of activity, whereas others send messages in more consistent intervals. For example, [22] observed that opinion spammers are rarely long-term users of any website, whereas genuine reviewers typically are. The posting of reviews over a long and short period of time indicates regular and suspicious activity, respectively [44], [46]. Therefore, this study employed review gap as a behavioral feature for identifying fake reviews and review spammers.

C. HLR MODEL

Hierarchical modeling can account for the hierarchical structure of data sets consequent to unobserved heterogeneity when individual observations are nested in some factors at a higher level of data structure, which can lead to dependency across observations [51], [52]. Consider the relationship between reviews and reviewers. Reviews written by a specific reviewer form a group with high homogeneity; in other words, the reviews in the group are not completely independent of each other. This is due to the inherent differences between the writing styles of that reviewer and other reviewers; reviews written by a specific reviewer are linked to that reviewer [17]. The nesting structure leads to an inadequacy for using single logistic regression model for predicting due to the violation of the assumption of independence of the residuals and the indetermination or variability among reviewers [18]. By contrast, HLR models can determine how a covariate measured at different levels of the hierarchy influences the response variable. This is accomplished by permitting group characteristics at higher levels of data structure to be involved in modeling individual outcomes [53].

We divided the HLR model into two levels, in which reviews are at the lower level (level 1) and are each nested within a certain reviewer (level 2). The first level is expressed as

$$\log\left(\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)}\right)=\beta_{0j}+\beta_{1j}X_{1ij}+\ldots+\beta_{qj}X_{qij} \quad (1)$$

where $\log\left(\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)}\right)$ is the log function of the odds. The odd is the probability that individual review i written by reviewer j is fake (denoted $Y_{ij}=1$) divided by the probability that individual review i written by reviewer j is genuine (denoted $Y_{ij}=0$). $X_{1ij}, X_{2ij}, \ldots, X_{qij}$ representing the linguistic features of reviews, are the predictors. β_{0j} is the intercept, and $\beta_{1j}, \beta_{2j}, \ldots, \beta_{qj}$ are the coefficients corresponding to the predictors $X_{1ij}, X_{2ij}, \ldots, X_{qij}$. Each coefficient captures the average effect of a level-1 predictor on log($odds$), becoming

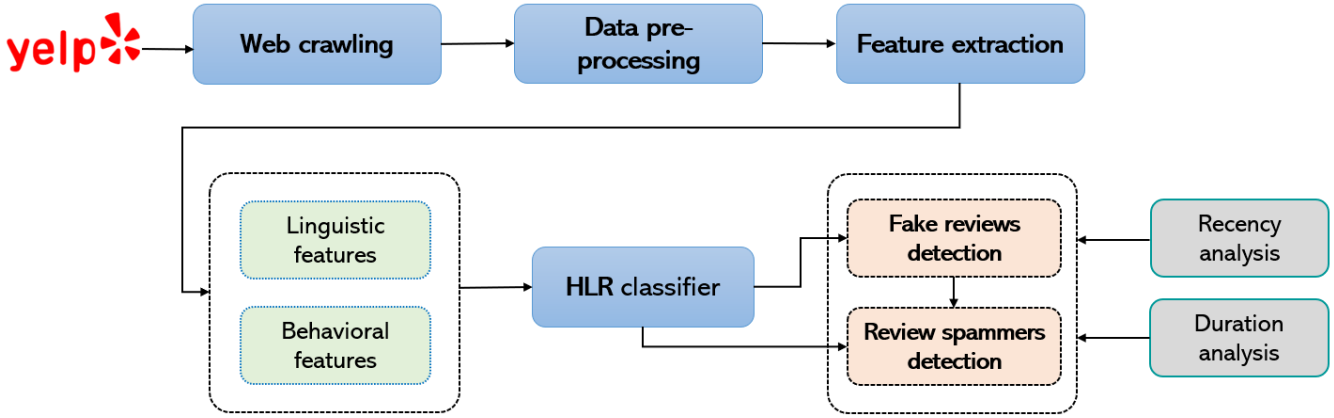

FIGURE 1. The system framework of the proposed HLR model.

the odds ratio OR when raised to the exponent, as in $OR_1 = \exp \beta_{1j}$. The (OR) is the multiplicative factor by which the predicted probability of an event occurring rather than not occurring $\left(\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)}\right)$ changes for a one-unit increase in the predictor variable X_{ij}.

At level 2, we assume that $\beta_{0j}, \ldots, \beta_{qj}$ depends on the unobserved factors specific to the jth reviewer. Thus,

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}Z_{1j} + \ldots + \gamma_{0k}Z_{kj} + u_{0j},\\ \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_{1j} + \ldots + \gamma_{1k}Z_{kj} + u_{1j}\\ &\ldots,\\ \beta_{qj} &= \gamma_{q0} + \gamma_{1q}Z_{1j} + \ldots + \gamma_{qk}Z_{kj} + u_{qj} \qquad (2)\end{aligned}$$

where u denotes the macro error, assumed to have a normal distribution $N(0, \delta^2)$; $\gamma_{00}, \ldots, \gamma_{q0}$ are the fixed effects; and $\gamma_{01}, \ldots, \gamma_{qk}$ are the coefficients associated with the behavioral features $Z_{1j}, \ldots, Z_{kj}$ of each reviewer. These coefficients represent the average effect of a level-2 variable on the log$(odds)$ and become the OR when raised to the exponent, as is the case with the coefficients in level 1. Equations (1) and (2) define a multilevel model that can equivalently be written as a single equation by substituting (2) into (1):

$$\begin{aligned}&\log\left(\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1}\right)\\ &= \gamma_{00} + \gamma_{01}Z_{1j} + \ldots \gamma_{0k}Z_{kj} + u_{0j} + (\gamma_{10}\\ &\quad +\gamma_{11}Z_{1j} + \ldots + \gamma_{1k}Z_{kj} + u_{1j})X_{1ij} + \ldots\\ &\quad +(\gamma_{q0} + \gamma_{1q}Z_{1j} + \ldots + \gamma_{qk}Z_{kj} + u_{qj})X_{qij} \qquad (3)\end{aligned}$$

The presence of macro error terms in (2) makes (3) a mixed model. If the macro errors are suppressed, (3) becomes a fixed-effect specification, and its estimation poses no particular problem. In this instance, eliminating the macro errors would be undesirable because we are unable to specify (even in principle) all the determinants of the within-reviewer coefficients. Our purpose was thus to incorporate this fundamental aspect of substantive formulation into an appropriate estimation procedure, following the recommendations of [54].

III. RESEARCH METHOD

Fig. 1 presents the system framework of the proposed HLR model. The model construction procedure was functionally divided into three major tasks. The first step involved retrieving and preprocessing a data set of reviews, including the content of the reviews. In the second step, the linguistic features of reviews and the behavioral features of reviewers were extracted to obtain model inputs. The third step entailed processing the two-level HLR model to classify fake reviews, in which level 1 represented review characteristics and level 2 represented behavioral features. This was followed by setting a threshold for review spammer detection. Recency and duration analyses were conducted to investigate the effectiveness of the proposed framework. Each step is explained in detail as follows.

A. DATA COLLECTION

We selected a labeled data set retrieved from the Yelp website. The data were originally collected by [46]. This data set comprised reviews written by 16 935 reviewers of 121 restaurants from 2004 to 2012 as well as corresponding information on the reviewers. From this initial data set, linguistic and behavioral features were extracted and then stored in the Review database.

B. DATA PREPROCESSING

Data preprocessing involved formalizing and structuring the review content in preparation for analysis. We followed the preprocessing procedure suggested in [55]. First, irrelevant characters, words, and other elements, such as HTML, tags, URLs, and punctuation marks, were eliminated. Next, the pronouns were replaced by corresponding nouns. Subsequently, the reviews were split into sentences in accordance with punctuation marks such as commas, semicolons, exclamation marks, and question marks. Finally, we applied POS

functions to the tokenized words and then removed the stop words.

C. FEATURE EXTRACTION

1) EXTRACTION OF LINGUISTIC FEATURES

As mentioned, each individual's linguistic style is unique and persists across multiple pieces of text. Herein, five linguistic aspects widely applied in the literature—LIWC, POS, readability, credibility, and evidentiality—were employed to determine review quality. The average reviewer tends to write more than one review in an online community; we thus had to aggregate the reviews posted by each reviewer. The extraction of linguistic features is explained as follows.

a: LIWC FACTORS

To analyze language use, we adopted the LIWC approach, which was developed by [24]. Four reliable LIWC factors—Immediacy, Making Distinctions, The Social Past, and Rationalization—were used to aid fake review detection. These factors comprised 11 subcategories in the linguistic dimension, namely affective processes, cognitive processes, negations, pronouns, quantifiers, social words, tentative words, word count, family-related words, leisure-related words, and words longer than six letters. Given a review, each LIWC factor was estimated for the kth category of LIWC of the jth review from the ith reviewer:

$$LIWC_{i,j,k} = \frac{wc_{i,j,k}}{w_{i,j}} \tag{4}$$

b: READABILITY

Various formulas for calculating the readability of a text have been developed over the past 80 years [27]. We conducted the Flesch reading ease test [56], a reliable, widely used measure [27], to calculate the readability score of each review, with higher scores indicating higher readability. The formula for the test [56] is

$$Readability_{i,j} = 206.835 - 1.015(\frac{w_{i,j}}{s_{i,j}}) - 84.6(\frac{sb_{i,j}}{w_{i,j}}) \tag{5}$$

where $sb_{i,j}$ is the number of syllables of each word of the jth review written by the ith reviewer and $s_{i,j}$ is the total sentences of jth review written by the ith reviewer.

c: CREDIBILITY

Credibility was determined on the basis of four indicators from the framework proposed by [57], namely capitalization, emoticons, shouting, and misspelling. As mentioned, appropriate use of capitalization represents a proper linguistic style, contributing to a sense of credibility. Regarding the emoticon indicator, the overuse of Western emoticons [e.g., :-) and :-D] reflects a less credible linguistic style [58], [59]. Writing in all capitals conveys a "shouting" tone, which is indicative of low credibility. Given that credible reviewers should be able to write with proper spelling, the more spelling errors were present in a text, the less credible we considered the text. Given a review, each kth indicator of credibility of the jth review written by the ith reviewer was estimated using the following pattern:

$$Credibility_{i,j,k} = \frac{x_{i,j,k}}{z_{i,j}} * 100\% \tag{6}$$

where $x_{i,j,k}$ is a parameter of the number of sentences beginning with a capitalized word and $z_{i,j}$ is the total number of sentences. Regarding the other indicators, $x_{i,j,k}$ refers to the number of words belonging to the kth category and $z_{i,j}$ represents the total word count of the jth review.

d: POS TAGGING

The literature on computational linguistics demonstrates that the frequency distribution of POS tags in a text is often dependent on the genre of the text [60]. Thus, we computed a feature POS distribution, which has been used in [10], to aid fake review detection. Four POS tagging categories were considered: verbs, adverbs, adjectives, and superlatives. The kth POS category of the jth review written by the ith reviewer was calculated as follows:

$$POS_{i,j,k} = \frac{wc_{i,j,k}}{z_{i,j}} \tag{7}$$

e: EVIDENTIALITY

Evidentiality, which is based on a hierarchy, forms a continuum from high to low. Various hierarchical schemes have been proposed. We employed the evidentiality categories proposed in [37] and classified them as representing high or low evidentiality (Table 1). The evidentiality score corresponding to the kth category of the jth review written by the ith reviewer is defined as

$$Evidentiality_{i,j,k} = \frac{wc_{i,j,k}}{w_{i,j}} \tag{8}$$

where $wc_{i,j,k}$ is the number of words belonging to the kth category corresponding to each feature of the jth review written by the ith reviewer and $w_{i,j}$ is the total word count of that review.

2) REVIEWER FEATURE EXTRACTION

Reviewer behavior analysis is essential to the detection of fake reviews and review spammers. We used the review gap, life tenure, review count, rating entropy, and rating deviation to determine review quality. Reviewer features were extracted as follows.

a: REVIEW GAP

The review gap was calculated as the time difference between the posting of two consecutive reviews written by a given reviewer. If a reviewer posts frequently, the review gap is extremely low. The equation for calculating the review gap (in days), presented in [50], is

$$Gap_i = \frac{1}{n_i - 1} \sum_{j=2}^{n_i} (t_{i,j} - t_{i,j-1}) \tag{9}$$

TABLE 1. Categories and items of evidentiality.

Category	Low evidentiality	High evidentiality
Attributive/ modal adverb	maybe, personally, perhaps, possibly, presumably seemingly, probably	certainly, sure, of course, definitely, absolutely, undoubtedly, clearly, obviously, apparently, really, always
Lexical verb	seem, think, sound, remember, observe, doubt, wish, wonder, infer, assume, forecast, fell, heard, hearsay	report, certain, believe, see
Auxiliary verb	ought, should, would, could, can, may, might	must
Epistemic adjective	definite	possible, likely, unlikely, probable, positive, potential, not sure, doubtful

where Gap_i corresponds to the review gap of the ith user, n_i is the number of reviews written by the ith user, and $t_{i,j}$ corresponds to the time stamp of the jth review posted by user i.

b: LIFE TENURE

Life tenure was calculated on the basis of an equation proposed in [50]:

$$Life_i = t_{i,n_i} - t_{i,0} \tag{10}$$

where $t_{i,0}$ is the time stamp of the first review written by the ith user and t_{i,n_i} is the time stamp of the jth review written by the ith reviewer.

c: RATING ENTROPY

Rating entropy was calculated on the basis of the entropy theory advanced by [43], in which the rating entropy of the ith reviewer is expressed as

$$Entropy_i = -\sum_{g=1}^{k} p_{i,g} log(p_{i,g}) \tag{11}$$

where $p_{i,g}$ is the probability of the ith reviewer giving a review score of g and k is the number of discrete rating scores that can be given by a reviewer.

d: RATING DEVIATION

Following [61] and [50], we computed the mean absolute deviation of each reviewer from the average rating of all restaurants reviewed by that reviewer to aid fake reviewer detection:

$$Deviation_i = \frac{1}{n_i} \sum_{j=1}^{n_i} |r_{i,j} - \mu_{h(j)}| \tag{12}$$

where $Deviation_i$ corresponds to the rating deviation of the ith reviewer, n_i is the number of reviews written by the ith reviewer, $r_{i,j}$ is the rating score given by the ith user for restaurant h_j in their jth review, and $\mu_{h(j)}$ is the mean rating of this restaurant.

3) HLR PROCEDURE

The linguistic and behavioral features were subjected to HLR as presented in [54]. This involved three crucial steps: 1) Running the model without predictors (i.e., constructing an empty model), 2) running the model with level 1 and level 2 predictors (i.e., constructing an intermediate model), and 3) constructing a final model by adding intra-level interactions.

The first step aims to confirm whether the data set has a nested structure. To calculate the intra-class correlation coefficient (ICC), an empty model (i.e., a model with no predictors) must be constructed. The ICC can be used to decompose the outcome variation into within-cluster and between-cluster variation [52], [54]. Furthermore, the ICC is a positive value between 0 and 1 and quantifies the proportion of between-cluster variation to the total outcome variation [62]. Depending on whether the ICC value conforms to the [0, 0.059], [0.059, 0.138], or [0.138, 1] interval, the degree of between-group heterogeneity is categorized as low, moderate, and high, respectively [52]. Moderate or high heterogeneity indicates that a data set has a nested structure and is thus suitable for HLR application. The ICC was calculated using (13), where $var(u_{0j})$ is the random intercept variance and $(\pi^2/3) \approx 3.29$ refers to the level-1 variance component in the standard logistic distribution [54]:

$$ICC = \frac{var(u_{0j})}{var(u_{0j}) + (\pi^2/3)} \tag{13}$$

After confirming that the data had a hierarchical data structure, we calculated the coefficient of the correlation between the independent variables on levels 1 and 2. If the correlation between any two variables had a large coefficient, we calculated the variance inflation factor (VIF) to determine whether multicollinearity was present between these variables. The VIF is a measure of how much the variance (the square of the estimate's standard deviation) in an estimated regression coefficient is increased because of collinearity [63]. Given that the effects of linguistic features depend on reviewer behavior, we constructed the model with level-1 and level-2 predictors (corresponding to the reviews and reviewers, respectively) to estimate the variation in the effect of linguistic features on the odds of a fake review from one

reviewer to another, since we expect the effect of linguistic features to depend on some reviewer's characteristics as presented in (1) and (2). Finally, as shown in (3), a synthesized model was run with level-1 predictors, level-2 predictors, and intra-level interactions to obtain the final model.

IV. RECENCY ANALYSIS

For experimentation, we used the data set collected by [46]. Yelp has a proprietary filtering algorithm for filtering out fake and suspicious reviews, which are presented in a list. Yelp also features recommended reviews considered to be genuine. Yelp's filter was reported to be highly accurate in an article published in [64]. For these reasons, we believe that the labeling of the Yelp data set is reliable and suitable for our purposes.

In this study, two experiments were conducted to evaluate the proposed model, namely recency and duration analyses. This section is dedicated to the former one.

A. DESCRIPTIVE ANALYSIS

In the recency analysis, the original Yelp data set was divided into three sub-data sets, designated as sub-data sets A, B, and C, containing each reviewer's 5, 30, and 50 most recent reviews, respectively. This process enabled accurate estimation of the regression coefficient [65], [66]. Table 2 displays the descriptive statistics of the sub-data sets. The first quartile of the mean number of words per reviewer (designated Q1) was defined as the middle number between the minimum value and the median, whereas the third quartile (designated Q3) was the middle value between the median and the maximum value. The obtained Q1 and Q3 values indicate slightly right-skewed distributions of the average number of words in a review. The discrepancy was greater under a greater number of reviews, suggesting that the more reviews written by a reviewer, the more information they wish to convey to other customers.

TABLE 2. Statistics of the sub-data sets for recency analysis.

Review description	5	30	50
Total reviews	69869	239860	310858
Average reviews per reviewer	4	14	18
Mean number of words per reviewer	392	1363	1786
Median of words per reviewer	391	862	874
Q1of words per reviewer	215	290	290
Q3 of words per reviewer	565	2262	2678

B. FAKE REVIEWS DETECTION

As mentioned in Section III, HLR was suitable for application to the three sub-data sets because of the presence of high heterogeneity. Sub-data sets A, B, and C had ICC values of 0.16, 0.28, and 0.31, respectively. An examination of the correlation coefficients and VIFs revealed no collinearity between variables in each sub-data set. As shown in Table 8 of Appendix, all the VIFs were smaller than 5. Next, the ORs of all predictors were calculated. When the OR of a feature was greater than 1, the greater the value, the more likely the review was classified as fake. Conversely, when the OR of a feature was less than 1, the greater the value, the less likely the review was classified as fake. Due to space limitations, the ORs and p values of all features and sub-data sets are presented in Table 9 of Appendix. The influence of each variable on the dependent variable differed between sub-data sets; however, these differences were not large.

For demonstration, Fig. 2 and Fig. 3 present the ORs of features with positive and negative effects for sub data-set A. Significant features are marked with an asterisk. Notably, when the rating entropy increased by one unit, the probability of the review being classified as fake was 6.132 times that of the probability of the review being classified as genuine. The rating deviation exhibited a similar trend, with an OR of 2.055. These results are consistent with those of [50]; this study also indicated that a higher rating entropy and rating deviation values are more likely to fake reviews. [2] empirically discovered that suspicious reviews tended to be more extreme than normal ones. To avoid being detected

FIGURE 2. ORs of features with positive effects for sub data-set A.

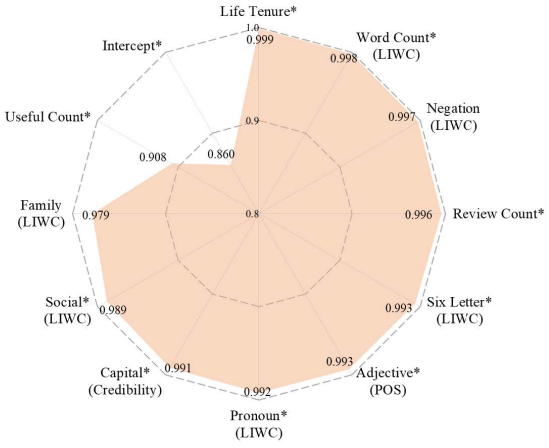

FIGURE 3. ORs of features with negative effects for sub data-set A.

TABLE 3. Hyperparameter values used in the experiments.

HLR			SVM		KNN	NB	RF	
Max integrations	Confidence level	C	Kernel	Gamma	Number of neighbors	Smoothing	Number of trees	Max depth
100	0.95	1	RBF	Scale	5	1.00E-09	900	None

and blocked by an online platform, a review spammer may have multiple online accounts. This corresponds to shorter life tenure, a longer review gap, and a lower review count. These inferences are in line with the results demonstrated in both figures. A fake review might involve more cognitive processes and be more effective than a genuine review, according to the findings of [23]. Review spammers usually use simple vocabulary and shorter words to enhance the readability of their reviews. Our finding was also consistent with the outcome of [15], which gave credit to the adoption of straightforward expressions for higher readability of fake reviews, attracting more review readers. Thus, fake reviews typically contain fewer words longer than six letters. Furthermore, they may contain improperly capitalized words and more misspellings, which indicate lower credibility. A less credible review may have a fewer useful count (e.g., total useful feedback that review receiving from readers).

To evaluate the detection performance of the proposed HLR model, the well-recognized machine learning algorithms, namely, NB, SVM, RF, and KNN, were chosen as the benchmarking models. They were implemented by Scikitlearn machine learning library, in which the default hyperparameter values were adopted for simplicity, as illustrated in Table 3. The performance evaluation was conducted with five-fold cross-validation in terms of accuracy, precision, recall, F1 score, and area under the receiving operator characteristic curve (AUC), as shown in Fig. 4. The HLR model had the most favorable performance; this was attributable to its consideration of the nested structure of the sub-data sets. It can be discovered that as the amount of data in the group increased, the detection performance declined. This can be explained by the increase of the variation of features extracted in reviews with the rising number of reviews.

C. REVIEW SPAMMERS DETECTION

Following [67], [68], we used the percentage of fake reviews to identify review spammers, setting a threshold in our experiment. When the percentage of fake reviews exceeded the threshold, a given reviewer was considered a review spammer.

We set the threshold to range from 10% to 90%. Through the detection of fake reviews, we further calculated the fake review rate of each reviewer. If the fake review rate was higher than the corresponding threshold, the reviewer was classified as a review spammer. Table 4 displays the performance indicators when using the HLR and LR models. Among most of the performance measures, HLR yielded more favorable detection results and the performance improved according

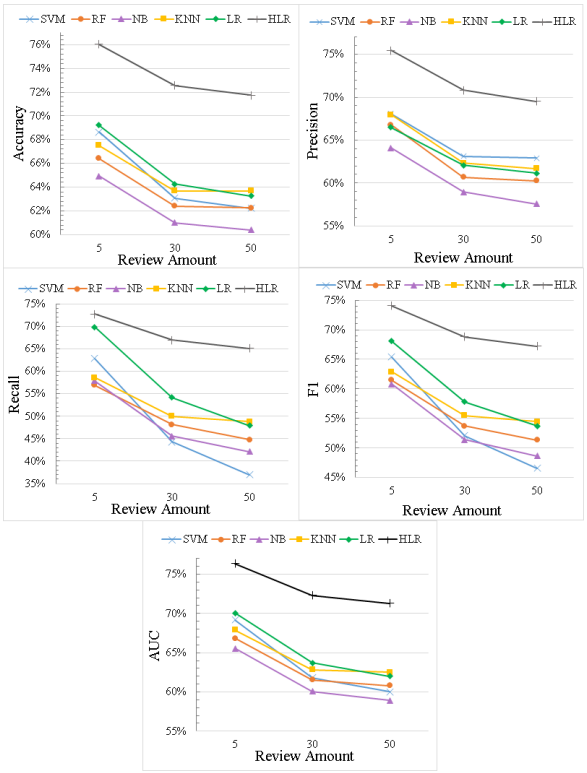

FIGURE 4. Results of fake reviews detection by all models for recency analysis.

to the rising amount of data from 10% to 50% whereas it decreased after exceeding the 50% threshold. The optimal results were observed when the threshold was set to 50%; thus, this threshold level was used in the performance comparison.

In Fig. 5, the HLR model obtained the most favorable detection outcomes. In all models, the detection accuracy increased upon the increasing number of reviews. This may be due to the fact that the more reviews written by a reviewer and included, the more accurately the determination of fake versus genuine could be made. In most of the detection results, the numerical differences between the HLR model and other models exceeded 20%. Overall, the results of fake review detection and review spammer detection both demonstrate that the HLR model is more suitable for application when the review data are hierarchical.

V. DURATION ANALYSIS

The purpose of the experiment was to explore the impact of different time intervals on fake review predictions.

TABLE 4. Results of review spammers detection using the HLR and LR models for recency analysis.

Threshold	Review Amount	HLR					LR				
		Accuracy	Precision	Recall	F1	AUC	Accuracy	Precision	Recall	F1	AUC
10%	5	61.5%	98.0%	60.6%	74.9%	69.3%	59.1%	94.2%	60.6%	73.8%	45.5%
	30	73.4%	98.6%	73.0%	83.9%	76.7%	75.3%	95.0%	78.0%	85.7%	50.4%
	50	**75.2%**	**98.6%**	**74.9%**	**85.1%**	**77.7%**	**77.3%**	**95.3%**	**80.1%**	**87.0%**	**52.3%**
20%	5	66.5%	97.7%	64.1%	77.4%	74.5%	60.1%	90.6%	61.8%	73.5%	52.6%
	30	76.0%	98.5%	74.4%	84.7%	82.5%	70.8%	91.1%	74.7%	82.1%	55.3%
	50	**77.7%**	**98.5%**	**76.2%**	**86.0%**	**83.3%**	**72.3%**	**91.3%**	**76.3%**	**83.2%**	**56.4%**
30%	5	69.7%	96.2%	65.2%	77.7%	77.1%	59.1%	85.1%	60.1%	70.5%	57.4%
	30	80.7%	97.7%	78.0%	86.7%	85.2%	68.3%	85.3%	73.7%	79.1%	59.5%
	50	**82.4%**	**97.7%**	**80.2%**	**88.1%**	**86.2%**	**69.2%**	**85.6%**	**74.7%**	**79.7%**	**60.1%**
40%	5	75.3%	91.6%	71.4%	80.2%	76.2%	61.5%	77.8%	63.4%	69.9%	59.8%
	30	86.4%	95.5%	84.6%	89.7%	87.2%	66.5%	77.4%	73.9%	75.6%	60.5%
	50	**87.9%**	**95.5%**	**87.0%**	**91.0%**	**88.3%**	**66.8%**	**77.6%**	**74.2%**	**75.9%**	**61.1%**
50%	5	79.0%	89.0%	74.2%	81.0%	78.1%	65.4%	73.3%	66.6%	69.8%	62.2%
	30	91.3%	92.7%	92.8%	92.7%	88.1%	**67.0%**	71.1%	**76.0%**	**73.5%**	**62.2%**
	50	**92.8%**	**92.8%**	**95.5%**	**94.1%**	**89.3%**	66.9%	**71.1%**	75.6%	73.3%	61.9%
60%	5	82.4%	**78.6%**	84.5%	81.5%	83.5%	**67.4%**	**62.0%**	73.8%	67.4%	**68.0%**
	30	**86.6%**	78.5%	97.3%	**86.9%**	**87.1%**	66.2%	59.8%	**79.0%**	**68.1%**	66.8%
	50	86.3%	77.4%	**98.7%**	86.8%	86.9%	65.5%	59.4%	77.1%	67.1%	66.0%
70%	5	**86.1%**	**74.4%**	92.3%	**82.4%**	**87.5%**	**70.2%**	**55.5%**	80.4%	**65.7%**	**72.7%**
	30	82.0%	66.5%	99.2%	79.6%	85.9%	67.3%	52.5%	**82.6%**	64.2%	70.9%
	50	80.9%	65.0%	**99.6%**	78.7%	85.0%	67.0%	52.2%	79.6%	63.0%	69.9%
80%	5	**85.2%**	**67.1%**	98.1%	**79.7%**	**90.1%**	**71.3%**	**51.0%**	**87.4%**	**64.4%**	**76.8%**
	30	79.8%	59.5%	99.8%	74.6%	85.7%	69.7%	49.4%	84.8%	62.4%	74.0%
	50	78.5%	58.0%	**99.9%**	73.4%	84.7%	69.5%	49.1%	80.9%	61.1%	72.5%
90%	5	**86.2%**	**65.9%**	99.6%	**79.3%**	**90.4%**	**74.5%**	**51.2%**	**87.6%**	**64.6%**	**78.7%**
	30	80.8%	58.1%	**99.8%**	73.4%	87.3%	73.5%	50.1%	81.9%	62.2%	76.4%
	50	80.0%	57.2%	99.8%	72.7%	86.6%	73.3%	49.8%	76.9%	60.5%	74.6%

TABLE 5. Descriptive statistics of all sub-data sets for various durations. (unit: month).

Review description	1	3	6	12	36	72
Total reviews	49250	88595	140310	235057	528721	694796
Average reviews per reviewer	3	6	9	14	31	41
Mean number of words per reviewer	294	504	796	1358	3167	4199
Median of words per reviewer	153	215	294	431	767	874
Q1of words per reviewer	95	122	144	166	274	290
Q3 of words per reviewer	314	492	703	1113	2307	2810

We performed duration analysis to identify notable observations occurring over 1, 3, 6, 12, 36, and 72 months.

A. DESCRIPTIVE STATISTICAL ANALYSIS

For the duration analysis, the Yelp data set was decomposed into six sub-data sets corresponding to the aforementioned six duration (designated sub-data sets 1–6). Table 5 displays the descriptive statistics of the sub-data sets. Regarding the difference between Q1 and Q3, the longer the duration, the more information reviewers wished to convey through their reviews.

B. FAKE REVIEWS DETECTION

We calculated ICC values to determine whether each sub-data set was suited for HLR application. The ICC of sub-data sets 1–6 was 0.16, 0.22, 0.25, 0.29, 0.36, and 0.39, respectively.

TABLE 6. ORs of all features for various durations. (Unit: month).

Feature		Time interval					
		1	3	6	12	36	72
Intercept		0.62**	0.45**	0.47**	0.57**	0.79**	0.77**
Readibility		1.00	1.00	1.00	1.00**	1.00**	1.00**
Credibility	Capital	0.52*	0.61*	0.66*	0.8	0.82*	0.63**
	Emoticon	0.50	0.58	0.39	0.32	0.26**	0.07**
	Shouting	1.31	1.98	1.56	1.83*	1.19	0.98
	Misspell	19.3	21.2	18.9	127.6**	81.2**	49.5**
Evidentiality	High	0.81	0.95	1.51	1.42	2.30**	2.46**
	Low	2.91	6.05*	8.85**	9.66**	22.2**	29.5**
LIWC	Affective	2.58*	2.61**	4.01**	5.24**	8.83**	12.2**
	Cognitive	1.68	1.46	0.94	0.80	0.60**	0.62**
	Negation	1.22	2.01	2.79*	1.68	2.47**	3.41**
	Pronoun	0.64	0.53*	0.43**	0.38**	0.34**	0.31**
	Quantifier	0.65	1.34	1.23	1.12	1.03	0.96
	Social	0.52*	0.42*	0.26**	0.18**	0.12**	0.11**
	Tentative	0.48	0.57	0.40*	0.45*	0.66*	0.60**
	Word Count	0.99**	0.99*	0.99*	1.00	1.00**	1.00**
	Leisure	1.71	1.26	0.98	1.29	1.72**	2.00**
	Family	0.24	0.09	0.02**	0.02**	0.00**	0.00**
	Six Letter	1.11	0.77	0.90	0.75	0.71**	0.65**
POS	Verb	1.99*	1.79*	1.63*	1.75**	1.51**	1.58**
	Adjective	0.68	0.87	0.56**	0.53**	0.46**	0.45**
	Adverb	1.09	1.59	1.31	1.33	1.58**	1.70**
	Superlative	3.15	1.51	0.92	1.25	1.12	0.90
Useful Count		0.90**	0.88**	0.90**	0.92**	0.92**	0.92**
Review Rating		1.01	1.03**	1.06**	1.06**	1.06**	1.06**
Review Gap		1.00**	1.00**	1.00**	1.00	0.99**	0.99**
Review Count		0.99**	0.99**	0.99**	0.99**	0.99**	0.99**
Rating Entropy		6.93**	8.02**	7.51**	6.16**	5.14**	5.77**
Rating Deviation		2.07**	2.27**	2.37**	2.41**	2.32**	2.32**
Life Tenure		0.99**	0.99**	0.99**	0.99**	0.99**	0.99**

Note: *: $p < 0.05$; **$p < 0.01$

TABLE 7. Results of review spammers detection using the HLR and LR algorithms for various durations.

Review Duration	Review count	HLR					LR				
		Accuracy	Precision	Recall	F1	AUC	Accuracy	Precision	Recall	F1	AUC
1-month	49,250	85.5%	**83.8%**	**82.8%**	**83.3%**	77.5%	77.4%	**72.0%**	**79.2%**	**75.4%**	71.8%
3-month	88,595	**85.8%**	82.3%	76.8%	79.4%	**81.7%**	**78.2%**	70.8%	66.5%	68.6%	**74.5%**
6-month	140,310	83.4%	79.2%	68.3%	73.4%	79.3%	76.2%	69.8%	50.9%	58.9%	70.2%
12-month	235,057	78.7%	75.0%	57.8%	65.3%	73.9%	71.5%	67.4%	34.8%	45.9%	63.4%
36-month	528,721	69.8%	67.4%	54.6%	60.3%	67.8%	61.4%	59.9%	25.3%	35.6%	56.7%
72-month	694,796	68.1%	65.7%	56.6%	60.8%	66.9%	59.9%	58.4%	28.5%	38.3%	56.5%

These values are considered sufficiently high for HLR application because they all exceed 0.138. An examination of the correlation coefficients and VIFs revealed no collinearity between variables in each sub-data set. As shown in Table 10 of Appendix, all the VIFs were smaller than 5. The ORs of all predictors were calculated and are presented in Table 6.

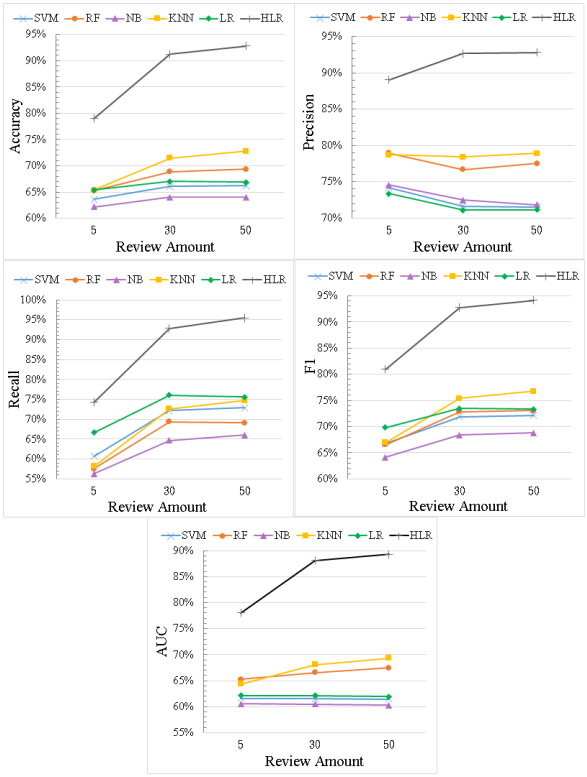

FIGURE 5. Results of review spammers detection by all models for recency analysis.

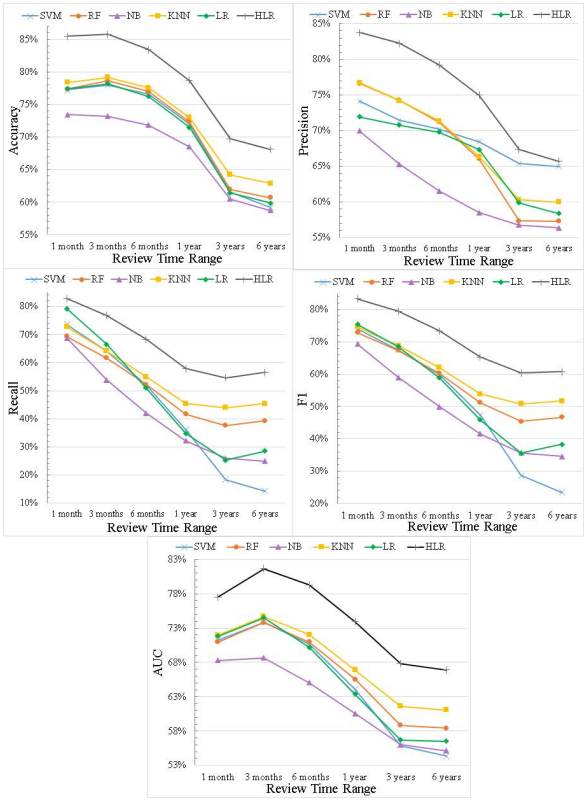

FIGURE 6. Results of fake reviews detection by all models for duration analysis.

Significant features are marked with asterisks. Similar to the results of the duration analysis, fake reviews had larger rating entropy, greater rating deviation, and a longer review gap but lower word count and useful count and exhibited improper use of capitalization. Furthermore, fake reviews were associated with shorter life tenure. The duration analysis revealed that affective processes and verbs positively affected fake review detection. Specifically, consistent with the findings of other studies, we determined that most spammers write imaginative reviews containing more pronouns, adverbs, and verbs, whereas genuine reviewers write informative reviews containing more adjectives and nouns [69], [70]. In line with those presented in [71], our results confirm that affective processes are a useful LIWC factor and contribute substantially to fake review detection. In addition, we validated the result with the work of [72] which also adopted the labelled data-set collected by [46]. The finding was consistent with ours in that extreme rating entropy, and rating deviations were the signs of fake reviews.

Finally, we evaluate the detection performance of the proposed model and compared to other machine learning algorithms. Consistent with the recency analysis findings, the HLR model outperformed the LR, SVM, RF, NB, and KNN models. Fig. 6 shows that in most of the detection results, the numerical differences between the HLR model and other models exceeded 10%. Moreover, for all models, detection accuracy decreased as the duration lengthened due to the increase of the variation in each sub-data set. Overall, the results of fake review detection and review spammer detection both demonstrate that the HLR model is more suitable for application when the review data are hierarchical.

C. REVIEW SPAMMERS DETECTION

As in the recency analysis, we set the review spammer detection threshold to range from 10% to 90%. Table 7 presents a comparison of the detection performance of the HLR and LR models. The optimal results were achieved under a threshold 50% and a period of 72 months (6 years). All performance indexes exceeded %. Notably, the recall reached 99%, indicating that almost all reviewers who were identified as review spammers were actually review spammers.

The threshold values used for the performance comparison are shown in Fig. 7. HLR achieved the most favorable detection outcomes. As in the recency analysis, detection performance improved as the duration lengthened. Overall, the results confirm the premise that the more reviews written by a given reviewer and included, the more accurately it can be determined whether that reviewer is a review spammer. In sum, both the recency and duration analyses revealed that HLR is the most suitable approach for handing the nesting present in most review data.

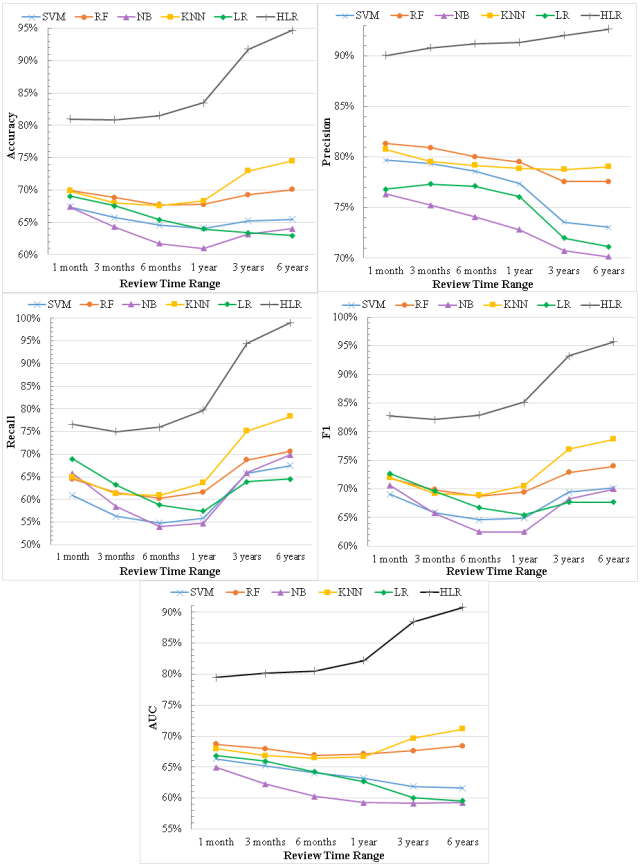

FIGURE 7. Results of review spammers detection by all models for duration analysis.

VI. CONCLUSION AND FUTURE DIRECTIONS

Because of the vast quantity of information available on online platforms, identifying credible reviews and reviewers, whose opinions consumers consider when making purchasing decisions, is essential. Studies on fake review detection have ignored the nested association between reviews and reviewers; this can undermine detection performance. To overcome this major shortcoming, we employed HLR to determine how much variance is ascribable to a reviewer and to each individual review on the basis of their nested connection. Recency and duration analyses were performed to investigate the role of linguistic and behavioral features in fake review detection and review spammer detection by considering the number of reviews and time stamps, respectively. Subsequently, a comparison of the detection performance of the HLR model and LR, SVM, RF, NB, and KNN models was undertaken.

The HLR model had the most favorable performance overall, demonstrating that the hierarchical review–reviewer relationship contributes crucially to detection accuracy and thus must not be disregarded. Fake reviews tended to have greater rating deviations and scoring deviations and to contain more emoticons and leisure-related words. Moreover, fake reviews involved more cognitive processes and corresponded to a longer review gap, shorter life tenure, a lower word count, a lower review count, more words longer than six letters, and more adjectives and pronouns. In addition, they were less likely to feature correct capitalization and had lower sociality and usefulness. In the recency analysis, the detection accuracy was maximized when the model was applied to sub-data set A (i.e., the sub-data set containing reviewers' five most recent reviews). In the duration analysis, the detection accuracy was highest when the reviews in one month of a reviewer adopted to HLR model. The results suggest that we should examine as many of a given reviewer's reviews as possible to optimize review spammer detection. Detection accuracy was greater in the duration analysis than in the recency analysis, indicating that the duration of past reviews has an effect on detection performance.

TABLE 8. VIFs of features in sub-data sets A, B, and C.

Feature		Review amount		
		5	30	50
Readibility		2.281	2.234	2.223
Credibility	Capital	1.437	1.430	1.418
	Emoticon	1.011	1.011	1.012
	Shouting	1.039	1.037	1.039
	Misspell	1.017	1.015	1.017
Evidentiality	High	1.241	1.239	1.235
	Low	1.166	1.171	1.172
LIWC	Affective	1.948	1.807	1.767
	Cognitive	2.295	2.354	2.372
	Negation	1.330	1.339	1.342
	Pronoun	1.710	1.701	1.702
	Quantifier	1.122	1.130	1.131
	Social	1.371	1.348	1.344
	Tentative	1.549	1.567	1.574
	Word Count	1.472	1.447	1.438
	Leisure	1.061	1.066	1.066
	Family	1.077	1.071	1.069
	Six Letter	2.239	2.162	2.144
POS	Verb	1.442	1.442	1.447
	Adjective	1.637	1.545	1.522
	Adverb	1.118	1.118	1.117
	Superlative	1.081	1.074	1.071
Useful Count		1.058	1.082	1.090
Review Rating		1.294	1.215	1.200
Review Gap		1.413	1.383	1.359
Review Count		1.361	1.356	1.357
Rating Entropy		1.437	1.298	1.283
Rating Deviation		1.180	1.149	1.156
Life Tenure		1.803	1.524	1.457

TABLE 9. ORs of features in sub-data sets A, B, and C.

Feature		Review Amount					
		5		30		50	
		odds ratio	p value	odds ratio	p value	odds ratio	p value
Intercept		0.860	0.000	0.835	0.000	0.828	0.000
Readibility		1.002	0.155	1.004	0.000	1.004	0.000
Credibility	Misspell	1.045	0.131	1.056	0.000	1.051	0.000
	shouting	1.007	0.101	1.002	0.355	1.002	0.261
	Emoticon	1.001	0.949	0.989	0.077	0.987	0.015
	Capital	0.991	0.000	0.993	0.000	0.993	0.000
Evidentiality	High	1.000	0.950	1.007	0.092	1.008	0.030
	Low	1.004	0.588	1.021	0.000	1.023	0.000
LIWC	Leisure	1.012	0.003	1.015	0.000	1.014	0.000
	Affective	1.015	0.000	1.019	0.000	1.020	0.000
	Negation	0.997	0.622	1.006	0.070	1.007	0.011
	Word Count	0.998	0.000	1.000	0.162	1.000	0.451
	Pronoun	0.992	0.003	0.988	0.000	0.988	0.000
	Family	0.979	0.110	0.951	0.000	0.945	0.000
	Six Letter	0.993	0.005	0.996	0.002	0.997	0.008
	Social	0.989	0.000	0.982	0.000	0.981	0.000
	Tentative	1.001	0.827	1.000	0.901	1.000	0.908
	Cognitive	1.007	0.041	0.999	0.377	0.997	0.056
	Quantifier	1.003	0.514	1.002	0.393	1.003	0.248
POS	Verb	1.001	0.613	1.001	0.386	1.001	0.407
	Adverb	1.002	0.433	1.005	0.003	1.005	0.000
	Superlative	1.008	0.289	1.003	0.522	1.001	0.711
	Adjective	0.993	0.007	0.993	0.000	0.994	0.000
Review Rating		1.001	0.906	1.024	0.000	1.027	0.000
Useful Count		0.908	0.000	0.875	0.000	0.875	0.000
Review Count		0.996	0.000	0.996	0.000	0.997	0.000
Life Tenure		0.999	0.000	0.999	0.000	1.000	0.000
Review Gap		1.002	0.000	1.000	0.619	0.999	0.004
Rating Entropy		6.132	0.000	6.083	0.000	5.878	0.000
Rating Deviation		2.055	0.000	2.310	0.000	2.315	0.000

A. THEORETICAL AND MANAGERIAL CONTRIBUTIONS

This main theoretical contribution of this study is that it proposes a new approach for detecting fake reviews by considering both the linguistic and behavioral aspects of review data. The proposed model outperformed various machine learning techniques because it has the ability to handle nested data. Notably, it can also be applied to identification of review spammers as a premodule of quality. Furthermore, we determined which linguistic style and behavioral features are important in detecting fake reviews. The findings serve as a reference for scholars and stakeholders alike with regard to understanding which features of review language and reviewer behavior most strongly affect review quality.

The present study has valuable managerial implications; the identification of fake reviews and reviewers from stylistic and behavioral perspectives enables potential consumers to avoid untrustworthy information and find genuine reviews on which to base their purchasing decisions. Specifically, our model can help manufacturers and retailers recognize review spammers who spread deceptive information and issue warnings in opinion-sharing communities accordingly. Moreover, companies can enlist the assistance of credible reviewers to support marketing campaigns. For example, during the life

TABLE 10. VIFs of all features for various duration.

Feature		Review Amount					
		1 month	3 month	6 month	1 year	3 year	6 year
Readibility		2.379	2.283	2.246	2.283	2.145	2.114
Credibility	Capital	1.504	1.426	1.422	1.419	1.387	1.367
	Emoticon	1.013	1.014	1.014	1.013	1.014	1.015
	Shouting	1.043	1.039	1.039	1.041	1.041	1.047
	Misspell	1.013	1.012	1.014	1.018	1.019	1.022
Evidentiality	High	1.280	1.265	1.252	1.246	1.229	1.222
	Low	1.161	1.165	1.164	1.167	1.173	1.175
LIWC	Affective	1.969	1.888	1.853	1.812	1.726	1.679
	Cognitive	2.313	2.346	2.348	2.365	2.392	2.401
	Negation	1.317	1.335	1.348	1.344	1.347	1.343
	Pronoun	1.710	1.696	1.698	1.714	1.717	1.717
	Quantifier	1.126	1.128	1.134	1.133	1.134	1.134
	Social	1.370	1.351	1.344	1.341	1.339	1.334
	Tentative	1.553	1.571	1.573	1.582	1.592	1.601
	Word Count	1.457	1.452	1.455	1.445	1.434	1.415
	Leisure	1.056	1.060	1.061	1.063	1.064	1.064
	Family	1.078	1.073	1.072	1.069	1.067	1.066
	Six Letter	2.428	2.312	2.267	2.262	2.130	2.092
POS	Verb	1.455	1.462	1.466	1.483	1.484	1.490
	Adjective	1.613	1.580	1.551	1.524	1.485	1.465
	Adverb	1.115	1.117	1.121	1.124	1.125	1.127
	Superlative	1.071	1.069	1.067	1.061	1.058	1.058
Useful Count		1.072	1.135	1.164	1.184	1.204	1.209
Review Rating		1.253	1.224	1.206	1.190	1.171	1.161
Review Gap		1.301	1.268	1.259	1.263	1.331	1.331
Review Count		1.454	1.549	1.532	1.484	1.553	1.556
Rating Entropy		1.369	1.319	1.284	1.250	1.246	1.236
Rating Deviation		1.196	1.189	1.188	1.187	1.228	1.245
Life Tenure		1.809	1.716	1.608	1.452	1.434	1.414

cycle of a product, manufacturers should encourage genuine reviewers to share their positive experiences with that product in advertisements. This type of campaign can persuade consumers to purchase the product, thus increasing sales. In addition, firms should pay attention to genuine negative reviews, such as suggestions for improving their products or services. Notably, our approach can be used on all types of website

B. LIMITATIONS AND FUTURE DIRECTIONS

This study has some limitations. First, the present data must be hierarchical in nature, containing information on both reviews and reviewers. Second, to identify the stable characteristics of reviewers, we examined review count as a behavioral feature and thus did not consider sentiment (because content words may vary substantially between topics). Third, we did not take into account other linguistic features, such as the feature of social networks and relationships (e.g., friendships) among platform users. The inclusion of such characteristics into future studies is expected to increase detection accuracy. Finally, we did not consider the emerging problem of spammer groups; this should be investigated in future studies.

APPENDIX

See Tables 8–10.

REFERENCES

[1] C. Xu, J. Zhang, K. Chang, and C. Long, "Uncovering collusive spammers in Chinese review websites," in *Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2013, pp. 979–988.

[2] M. Luca and G. Zervas, "Fake it till you make it: Reputation, competition, and yelp review fraud," *Manage. Sci.*, vol. 62, no. 12, pp. 3412–3427, 2016.
[3] L. de Vries, S. Gensler, and P. S. H. Leeflang, "Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing," *J. Interact. Marketing*, vol. 26, no. 2, pp. 83–91, May 2012.
[4] S. Feng, L. Xing, A. Gogar, and Y. Choi, "Distributional footprints of deceptive product reviews," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 6, 2012, pp. 98–105.
[5] M.-Y. Day, C.-C. Wang, C.-C. Chen, and S.-C. Yang, "Exploring review spammers by review similarity: A case of fake review in Taiwan," in *Proc. 3rd Int. Conf. Electron. Softw. Sci. (ICESS)*, 2017, p. 166.
[6] D. Mayzlin, Y. Dover, and J. Chevalier, "Promotional reviews: An empirical investigation of online review manipulation," *Amer. Econ. Rev.*, vol. 104, no. 8, pp. 55–2421, 2014.
[7] N. Jindal and B. Liu, "Review spam detection," in *Proc. 16th Int. Conf. World Wide Web (WWW)*, 2007, pp. 1189–1190.
[8] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proc. Int. Conf. Web Search Web Data Mining (WSDM)*, 2008, pp. 219–230.
[9] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2010, pp. 939–948.
[10] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," 2011, *arXiv:1107.4557*.
[11] R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," *Inf. Process. Manage.*, vol. 56, no. 4, pp. 1234–1244, Jul. 2019.
[12] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making," *Ind. Marketing Manage.*, vol. 90, pp. 523–537, Oct. 2020.
[13] J. Fontanarava, G. Pasi, and M. Viviani, "Feature analysis for fake review detection through supervised classification," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2017, pp. 658–666.
[14] I. Dematis, E. Karapistoli, and A. Vakali, "Fake review detection via exploitation of spam indicators and reviewer behavior characteristics," in *Proc. Int. Conf. Current Trends Theory Pract. Informat.* Cham, Switzerland: Springer, 2018, pp. 581–595.
[15] S. Banerjee and A. Y. K. Chua, "A study of manipulative and authentic negative reviews," in *Proc. 8th Int. Conf. Ubiquitous Inf. Manage. Commun. (ICUIMC)*, 2014, pp. 1–6.
[16] J. W. Pennebaker and T. C. Lay, "Language use and personality during crises: Analyses of mayor rudolph Giuliani's press conferences," *J. Res. Personality*, vol. 36, no. 3, pp. 271–282, Jun. 2002.
[17] S.-T. Li, T.-T. Pham, and H.-C. Chuang, "Do reviewers' words affect predicting their helpfulness ratings? Locating helpful reviewers by linguistics styles," *Inf. Manage.*, vol. 56, no. 1, pp. 28–38, Jan. 2019.
[18] P. Bressoux, *Modélisation Statistique Appliquée Aux Sciences Sociales*. Brussels, Belgium: De boeck Bruxelles, 2010.
[19] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2014, pp. 1566–1576.
[20] D. Liang, X. Liu, and H. Shen, "Detecting spam reviewers by combing reviewer feature and relationship," in *Proc. Int. Conf. Informative Cybern. Comput. Social Syst. (ICCSS)*, Oct. 2014, pp. 102–107.
[21] C. M. Aye and K. M. Oo, "Review spammer detection by using behaviors based scoring methods," in *Proc. Int. Conf. Adv. Eng. Technol.*, 2014, pp. 350–355.
[22] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in *Proc. 21st Int. Conf. World Wide Web (WWW)*, 2012, pp. 191–200.
[23] S. M. Ho and J. T. Hancock, "Computer-mediated deception: Collective language-action cues as stigmergic signals for computational intelligence," in *Proc. 51st Hawaii Int. Conf. Syst. Sci.*, 2018, pp. 1–10.
[24] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual difference," *J. Personality Social Psychol.*, vol. 77, no. 6, p. 1296, 1999.
[25] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," Univ. Texas Austin, Austin, TX, USA, Tech. Rep., 2015. [Online]. Available: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf?Sequence=3
[26] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *J. Big Data*, vol. 2, no. 1, pp. 1–24, Dec. 2015.
[27] W. H. DuBay, "The principles of readability," Impact Inf., Costa Mesa, CA, USA, Tech. Rep., 2004. [Online]. Available: https://files.eric.ed.gov/fulltext/ED490073.pdf
[28] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512, Oct. 2011.
[29] G. R. Klare, *Measurement of Readability*. Ames, IA, USA: Iowa State Univ. Press, 1963. [Online]. Available: https://www.amazon.com/Measurement-Readability-George-R-Klare/dp/B000LZOG7W
[30] B. L. Zakaluk and S. J. Samuels, *Readability: Its Past, Present, and Future*. ERIC, 1988.
[31] D. Zhang, L. Zhou, J. L. Kehoe, and I. Y. Kilic, "What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews," *J. Manage. Inf. Syst.*, vol. 33, no. 2, pp. 456–481, 2016.
[32] Y.-C. Ku, C.-P. Wei, and H.-W. Hsiao, "To whom should I listen? Finding reputable reviewers in opinion-sharing communities," *Decis. Support Syst.*, vol. 53, no. 3, pp. 534–542, 2012.
[33] V. L. Rubin and E. D. Liddy, "Assessing credibility of weblogs," in *Proc. AAAI Spring Symp., Comput. Approaches Analyzing Weblogs*, 2066, pp. 187–190.
[34] M. J. Metzger, "Making sense of credibility on the web: Models for evaluating online information and recommendations for future research," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 13, pp. 2078–2091, 2007.
[35] M. Bilal, M. Marjani, M. I. Lali, N. Malik, A. Gani, and I. A. T. Hashem, "Profiling users' behavior, and identifying important features of review 'Helpfulness,'" *IEEE Access*, vol. 8, pp. 77227–77244, 2020.
[36] W. L. Chafe and J. Nichols, *Evidentiality: The Linguistic Coding of Epistemology*. Norwood, NJ, USA: Praeger, 1986. [Online]. Available: https://www.amazon.com/Evidentiality-Linguistic-Epistemology-Discourse-Processes/dp/0893912034
[37] Q. Su, C.-R. Huang, and H. K. Chen, "Evidentiality for text trustworthiness detection," in *Proc. Workshop NLP Linguistics, Finding Common Ground*, 2010, pp. 10–17.
[38] R. Jakobson, *Shifters and Verbal Categories*. Cambridge, MA, USA: Harvard Univ. Press, 1990.
[39] S. DeLancey, "The mirative and evidentiality," *J. Pragmatics*, vol. 33, no. 3, pp. 369–382, Mar. 2001.
[40] C. Tan, L. Lee, and B. Pang, "The effect of wording on message propagation: Topic- and author-controlled natural experiments on Twitter," 2014, *arXiv:1405.1438*.
[41] R. K. Roul, S. R. Asthana, M. Shah, and D. Parikh, "Detecting spam web pages using content and link-based techniques," *Sadhana*, vol. 41, no. 2, pp. 193–202, Feb. 2016.
[42] R. K. Dewang and A. K. Singh, "Identification of fake reviews using new set of lexical and syntactic features," in *Proc. 6th Int. Conf. Comput. Commun. Technol.*, 2015, pp. 115–119.
[43] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[44] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 632–640.
[45] Y. Lu, L. Zhang, Y. Xiao, and Y. Li, "Simultaneously detecting fake reviews and review spammers using factor graph model," in *Proc. 5th Annu. ACM Web Sci. Conf. (WebSci)*, 2013, pp. 225–233.
[46] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?" in *Proc. 7th Int. AAAI Conf. Weblogs Social Media*, 2013, pp. 409–418.
[47] L. Khansa, X. Ma, D. Liginlal, and S. S. Kim, "Understanding members' active participation in online question-and-answer communities: A theory and empirical analysis," *J. Manage. Inf. Syst.*, vol. 32, no. 2, pp. 162–203, Apr. 2015.
[48] M. Ma and R. Agarwal, "Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities," *Inf. Syst. Res.*, vol. 18, no. 1, pp. 42–67, Mar. 2007.
[49] P. B. Goes, M. Lin, and C. A. Yeung, "'Popularity effect' in user-generated content: Evidence from online product reviews," *Inf. Syst. Res.*, vol. 252, no. 2, pp. 222–238, 2014.
[50] N. Kumar, D. Venugopal, L. Qiu, and S. Kumar, "Detecting review manipulation on online platforms with hierarchical supervised learning," *J. Manage. Inf. Syst.*, vol. 35, no. 1, pp. 350–380, Jan. 2018.
[51] D.-G. Kim, Y. Lee, S. Washington, and K. Choi, "Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models," *Accident Anal. Prevention*, vol. 39, no. 1, pp. 125–134, 2007.

[52] L. M. O'Dwyer and C. E. Parker, "A primer for analyzing nested data: Multilevel modeling in SPSS using an example from a REL study. REL 2015-046," Regional Educ. Lab. Northeast Islands, Washington, DC, USA, Tech. Rep. REL 2015-046, 2014. [Online]. Available: https://files.eric.ed.gov/fulltext/ED551064.pdf
[53] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[54] N. Sommet and D. Morselli, "Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using stata, R, Mplus, and SPSS," *Int. Rev. Social Psychol.*, vol. 30, no. 1, pp. 203–218, Sep. 2017.
[55] S. Tirunillai and G. J. Tellis, "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation," *J. Marketing Res.*, vol. 51, no. 4, pp. 463–479, Aug. 2014.
[56] R. Flesch, "A new readability yardstick," *J. Appl. Psychol.*, vol. 32, no. 3, p. 221, 1948.
[57] W. Weerkamp and M. De Rijke, "Credibility improves topical blog post retrieval," in *Proc. ACL: HLT*, 2008, pp. 923–931.
[58] R. B. Harris and D. Paradice, "An investigation of the computer-mediated communication of emotions," *J. Appl. Sci. Res.*, vol. 3, no. 12, pp. 2081–2090, 2007.
[59] V. Griskevicius, N. J. Goldstein, C. R. Mortensen, J. M. Sundie, R. B. Cialdini, and D. T. Kenrick, "Fear and loving in las vegas: Evolution, emotion, and persuasion," *J. Marketing Res.*, vol. 46, no. 3, pp. 384–395, Jun. 2009.
[60] N. Silveira, T. Dozat, M.-C. De Marneffe, S. R. Bowman, M. Connor, J. Bauer, and C. D. Manning, "A gold standard dependency corpus for English," in *Proc. LREC*, 2014, pp. 2897–2904.
[61] B. Jiang, R. H. Cao, and B. Chen, "Detecting product review spammers using activity model," in *Proc. Int. Conf. Adv. Comput. Sci. Electron. Inf. (ICACSEI)*, 2013, pp. 650–653.
[62] G. Shieh, "Choosing the best index for the average score intraclass correlation coefficient," *Behav. Res. Methods*, vol. 48, no. 3, pp. 994–1003, Sep. 2016.
[63] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 112. New York, NY, USA: Springer, 2013.
[64] K. Weise, "A lie detector test for online reviewers," *Bloomberg Business Week*, Sep. 29, 2011.
[65] S. Ali, A. Ali, S. A. Khan, and S. Hussain, "Sufficient sample size and power in multilevel ordinal logistic regression models," *Comput. Math. Methods Med.*, vol. 2016, pp. 1–8, Sep. 2016.
[66] R. Moineddin, F. I. Matheson, and R. H. Glazier, "A simulation study of sample size for multilevel logistic regression models," *BMC Med. Res. Methodol.*, vol. 7, no. 1, pp. 1–10, Dec. 2007.
[67] S. Kim, H. Park, and G. Lebanon, "Fast spammer detection using structural rank," 2014, *arXiv:1407.7072*.
[68] J. Huang, T. Qian, G. He, M. Zhong, and Q. Peng, "Detecting professional spam reviewers," in *Proc. Int. Conf. Adv. Data Mining Appl.* Berlin, Germany: Springer, 2013, pp. 288–299.
[69] J. K. Rout, S. Singh, S. K. Jena, and S. Bakshi, "Deceptive review detection using labeled and unlabeled data," *Multimedia Tools Appl.*, vol. 76, no. 3, pp. 3187–3211, Feb. 2017.
[70] M. N. I. Ahsan, T. Nahian, A. A. Kafi, M. I. Hossain, and F. M. Shah, "Review spam detection using active learning," in *Proc. IEEE 7th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 2016, pp. 1–7.
[71] X. Wang, X. Zhang, C. Jiang, and H. Liu, "Identification of fake reviews using semantic and behavioral features," in *Proc. 4th Int. Conf. Inf. Manage. (ICIM)*, May 2018, pp. 92–97.
[72] J. Wang, H. Kan, F. Meng, Q. Mu, G. Shi, and X. Xiao, "Fake review detection based on multiple feature fusion and rolling collaborative training," *IEEE Access*, vol. 8, pp. 182625–182639, 2020.

THI-KIM-HIEN LE received the M.S. degree in management information systems from the Ho Chi Minh City University of Technology, Vietnam National University, Ho Chi Minh City, in 2014. She is currently pursuing the Ph.D. degree with the Institute of Information Management, National Cheng Kung University, Taiwan. She is also an Instructor with the School of Information Management, University of Economics and Law, Vietnam National University. Her research interests include business intelligence, data mining, text mining, and human–computer interaction.

YI-ZHEN LI received the M.S. degree in information management from the National Cheng Kung University, Taiwan, in 2020. She is specialized in artificial intelligence and text mining. She is currently a Software Engineer and her work focus on developing the software for semiconductor manufacturing.

SHENG-TUN LI received the Ph.D. degree in computer science from the University of Houston, University Park, TX, USA, in 1995. He is currently a Distinguished Professor with the Department of Industrial and Information Management and the Institute of Information Management, National Cheng Kung University, Taiwan. He is the author/coauthor of ten IT-related textbooks, including two translated, over 80 journal articles, and numerous conference papers. His work has been appeared in *Information & Management*, *Omega-The International Journal of Management Science*, IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, *Fuzzy Sets and Systems*, *Information Sciences*, *Journal of Information Science*, and *Technovation*. He is also a holder of one IT-related patent. His research interests include artificial intelligence, business intelligence, data mining, and text mining.