Hindawi
Mobile Information Systems
Volume 2023, Article ID 2565020, 11 pages
https://doi.org/10.1155/2023/2565020

Research Article

What Is the Internet Water Army? A Practical Feature-Based Detection of Large-Scale Fake Reviews

Bo Guo and Zhi-bin Jiang

School of Business and Management, Shanghai International Studies University, Shanghai 201600, China

Correspondence should be addressed to Zhi-bin Jiang; jiangzhibin@shisu.edu.cn

Received 16 September 2022; Revised 29 November 2022; Accepted 26 December 2022; Published 31 January 2023

Academic Editor: Manoj Kumar

Information is passed by word-of-mouth figures prominently when consumers evaluate products through reviews. However, severe logistical problems are caused by the internet's Water Army (i.e., literally people who are hired by individuals or organizations to compose false reviews), that flood the internet e-commerce websites. An array of internet e-commerce sites is flooded with inauthentic information, and false reviews are used to maliciously induce consumers to purchase specific products, that often contain some defects. Notwithstanding the fact that the internet Water Army first manifested in China, it can also exist in other countries. The rationale lies in the high profitability possible, in the minds of numerous organized underground paid poster groups, and in writing fake reviews to misinform consumers. It has become an increasingly daunting task to precisely spot the Water Army members, who often alter their writing style and posted content. In this paper, the authors devise a comprehensive set of features to characterize all users and compare the paid posters against the normal users on different dimensions; furthermore, an ensemble detection model equipped with seven disparate algorithms is put into place. Our model reached a score of 0.730 in the AUC measure, 0.691 in the $F1$ measure on the JD dataset, 0.926 in the AUC measure, and 0.871 in the $F1$ measure on the Amazon dataset, which outshines the measures in the existing research. The significance and contribution of this work are in advancing constructive solutions and recommendations for this major concern of the entire e-commerce industry.

1. Introduction

An official report by the China Internet Network Information Center (CNNIC) said that there are currently approximately 731 million internet users in China, which is approximately 53% of its total population [1]. Thanks to a huge pool of internet users, China's e-commerce industry has gained momentum over the past few years. The unprecedented development of e-commerce created not only profit, fame, and tax income but also fostered underground economic activities, which negatively affects this growing industry. For consumers who go shopping online, the reviews posted on a product detail page play an important purchase persuading power role, particularly for those who are hesitating among numerous choices. The studies in [2–4] all stressed the crucial influence that online reviews have on consumers. Paid posters analyzed eBay's feedback system and demonstrated the importance of meaningful feedback. The review pages, as a major source of persuasive information available to prospective customers, have been revealed to underpin a novel behind-the-scenes industry, namely, paid posters. The phrase, internet Water Army, has also recently gained popularity in this context [5]. The Water Army mainly provides the following two kinds of services to its customers:

(i) Promotion of a specific product, company, person, or message

(ii) Smears/slanders of counterparts who operate products or services of the same type

Figure 1 illustrates the typical business process of the Water Army. Another related concept is electronic spamming, the use of electronic messaging systems to send unsolicited messages [6]. The aforementioned business process

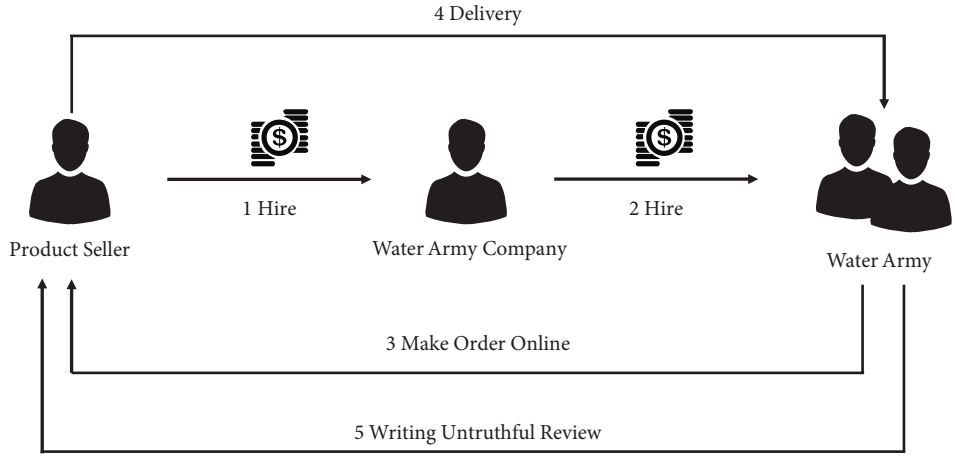

Figure 1: Typical business process of the internet Water Army.

leads to the conclusion that electronic spamming is a different concept than paid posters.

Spammers have the objective of distributing a large quantity of junk messages in a short period; furthermore, their focus is not on the content quality; instead, it is on the quantity of coverage. In contrast, the priority of paid posters lies in an increase in customers. Due to this, putting stress on the content and the number of posts is a requirement. A Water Army focuses on the content and quality of the reviews. They pose as average users or make rigid comments.

The offer of an online paid poster position is a desirable employment option for many netizens; among them, the majority is comprised of university students and the unemployed. The vast organized army of users of this sort "floods" the internet with purposeful reviews and articles. Online paid posters are put on the payroll of some public relationship (PR) companies and are asked to post reviews and articles on various online communities and websites. Companies are always interested in effective strategies to attract public attention to their products; in addition, this has evolved into a form of junk message in China's e-commerce industry. In a typical case provided in [5], a thread with blank content could receive as many as 300,621 replies and more than 7 million clicks in only two days. This level of the propensity for paid posters is widely seen in China but quite rare in other countries. The paid posters are quite well organized like an army; hence, they are called the Water Army. In recent years, some paid posters in China have even established companies (e.g., https://shuijun.co) to promote themselves. They have formed a perfect industrial chain that contains both the supply side (paid poster company) and demand side (PR company). There are quite a few signs that the Water Army industry has grown quite prosperous. In addition, a large prevalence of paid posters can also be seen in other countries [7].

However, the rise of the Water Army is detrimental to the fast-growing e-commerce industry. Online shopping consumers rely heavily on reviews to choose from numerous products that vary widely in quality. If the product detail pages are flooded by untruthful reviews, the consumers can be misled to buy those inferior products. From a long-term perspective, this is economically inefficient for the following two reasons. First, consumers' well-being could deteriorate due to low product quality. Second, this imperfect information situation could cause a market failure, in that the more efficient companies receive less market share.

Despite the fact that the Water Army exists as a severe problem, limited attention has been given to its related studies. Some scholars have probed into strategies that challenge the spammers; however, the tactics failed to face and struggle with the Water Army directly, given their differences from spammers. The following two questions remain to be addressed:

(1) How can a paid poster user be unmasked?

(2) How can we act to keep a tight rein on paid posters?

The greatest challenge faced in this research is that the paid posters will modify their own commentary style and deliberately imitate the writing style of others. The literature proposed solutions to this problem through the use of text similarity analysis [8], user group analysis [9], or the temporal features [10]. What they failed to deliver is completeness. The detection frameworks of the existing studies rest on a single factor, while certainly paid posters appear common in one aspect but alien in another. Shifting between multiple aspects shields the posters from the possibility of being ignored, which would undo their usefulness.

This study contributes to both practice and theoretical research in the following aspects [11]:

(1) We devise an extensive set of features to characterize an entire set of users and measure the paid posters in anticipation of common users.

(2) We use the text and meta-information of every review and the associated product's information, which is disregarded in the existing research.

(3) Seven classification algorithms are incorporated to establish an ensemble classification model.

(4) The primary attention of the previous studies is paid to the datasets from the US. In light of the universality and acuteness of the paid poster issue within

China, two datasets from the two countries are collected as a sample to probe the validity of our model. The two platforms chosen involve JD.com and Amazon.

(5) With reference to the analysis effect of the two datasets, the model is shown to hold a strong ability to differentiate and outshine the existing research.

The paper is organized as follows:

(1) Section 2 gives a survey of both the theoretical and practical studies on this topic, summarizing why the Water Army harms the e-commerce industry and how to detect paid posters.

(2) Section 3 presents a comprehensive model for paid poster detection.

(3) Section 4 evaluates the model on the JD dataset and the Amazon dataset. This section offers analytical insights into the accuracy and performance of the model.

(4) Section 5 summarizes the research content and suggests future research directions.

2. Related Work

A variety of detection approaches have been proposed since the pioneering work of [8]. Most studies used supervised learning algorithms, and certain behavioral features were formulated within those processes, including temporal patterns and textual features. The study of [8] investigated the reliability of online opinions within the sphere of product reviews and concluded that the false reviews greatly varied from traditional email spam. However, these studies mostly rested on twin reviews that do not exist at this moment as a result of strict webpage regulations. The study of [12] explored the lack of anticipation of reviews and advanced a private domain technological approach to identify today's professional Water Army society. The study of [10] focused on the inflammatory nature of reviews to locate the spammers. Upticks in review occurrences may originate from two sources, the instant favor of consumers as well as from spam attacks. The study of [13] studied the detection of false comments in online forums through text and sentiment analysis approaches.

Other studies have focused on the design of the textual and behavioral features. The studies of [14, 15] provided some insights into ways to characterize each person's writing style. The study of [16] explored the use of semantics in spam filtering by introducing a preprocessing word sense disambiguation step, which could detect the internal semantics of the spam messages. The study of [17] found that the readability of the review is more likely to impress consumers than the length of the review. The study of [18] demonstrated that semantic characteristics are more influential than other characteristics in affecting how many helpfulness votes reviews receive and that reviews with extreme opinions gained more support in comparison to those with ambivalent or unbiased opinions.

Highly advanced text mining and semantic analysis techniques have been used in this field, such as sentiment analysis and opinion extraction [19, 20]. Via the sentiment analysis channel, every review was deemed to be a note labeled with a sentiment (positive, negative, and neutral); afterwards, the challenge was addressed through the application of existing classification algorithms. The study of [21] used a text mining model and a semantic language model to solve the spammer detection problem. Semi-supervised methods were also used to levy the need for a large amount of training data [22]. With a focus on Dianping, China's comprehensive review hosting and rating application, the study of [23] devised an undivided classification algorithm following the employment of a model that collects both favorable and unlabeled samples. The study of [24] focused on how to use the characteristics of the relationship between people in social networks to detect the senders of false messages in the networks.

The key shortcoming of these studies was the incompleteness of their detection framework. Their method merely found the unexpected behaviors through unexpected rules from a specific aspect, instead of using a global and comprehensive rule to segment out the suspicious users. Some of the paid poster's actions are expected in one single regard, while they were uncommon in other respects. They could be overlooked in the detection process. For those studies that used text similarity as a key feature, because paid posters may alter their writing style easily, it is quite simple for paid posters to avoid being detected. These types of frameworks may fail soon after paid posters know the exact features. Some invariant characteristics of paid posters need to be discovered.

3. Detection Framework

This section presents a precise description of the framework, involving a common procedure of the e-commerce business, its feature system, the datum, and the classification algorithms. In light of the current research gap, the authors recommend an extensive scheme covering the following three prime regards:

Linguistic. These factors attend to the grammatical and emotional traits of the reviews posted by each consumer.

Behavioral. These factors chiefly draw on the metadata of the reviews, for instance, the post review time in making an order, together with the account profile.

Product. These factors consider the product information. Expressive styles may change effortlessly, while the brand is something that does not change. A disagreement between the product information and the contents of a review allows a revelation of differentiation abilities.

3.1. General Business Process of E-Commerce. For ordinary consumers, a typical online shopping experience process is shown in Figure 2.

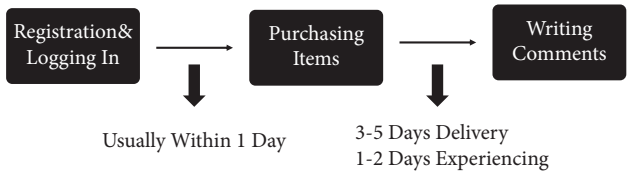

Figure 2: The online shopping process.

By analyzing the real user profile data and the review data, more information can be identified and applied to build the detection model. For instance, the user ought to register an account on the website that is used for the successive contacts from the website, and the username will be recorded when composing the reviews. After picking out the desired product, the user logs in to make an order to complete a valid order, since registration and logging in are mandatory. There is a delivery period, and the consumers themselves also need some time to take the delivery and evaluate the product quality. However, for paid posters, their behavior will somehow differ from that of normal users. In plain words, they may register several accounts in a batch or write reviews shortly before receiving the product, which will expose their true identity.

3.2. Feature Selections. The key notations used in this section are listed in Table 1.

3.2.1. Internal Text Similarity (ITS). The paid posters are highly motivated to reduce the time needed to compose the reviews. This explains why the same content may be present for various goods. Calculating the mean value of the text similarity level against the backdrop of all the reviews from one individual may serve as a feature targeting this phenomenon, and its definition is as follows. $\cos(c_{i,j}, c_{i,k})$represents the cosine similarity between two consumer reviews. The feature measures the text similarity within one user's reviews, for which this feature is referred to as internal.

$$\mathrm{ITS}_i = \frac{2}{n_i(n_i-1)} \sum_{j=1}^{n_i} \sum_{k\geq 1}^{n_i} \cos\left(c_{i,j}, c_{i,k}\right). \tag{1}$$

3.2.2. Comment Latency (CL). In light of the duration of a product delivery, a review always occurs several days after a purchase; the duration time is a key performance indicator for supply chain management in the e-commerce industry and is called click to deliver (C2D for short). In addition, a process is required to ensure that the goods are put into actual use by the purchaser. In this way, the quality of the product can be tested after it has been delivered. Consequently, a sensible billing postponement period exists in the interval between the review period and the payment receipt period. In contrast, the Water Army could present immediate reviews or quick reviews in the absence of a real delivery experience. This explains why we employ the review latency as a feature and refer to its measurement in the following equation:

Table 1: Symbol definition.

Notation	Definition
i	A user who made a comment
C_i	The set of reviews created by consumer i
n_i	The number of reviews created by consumer i
$\mathbf{C}$	The total set of all the reviews
$c_{i,j}$	A specific review j created by consumer i
$c_{i,j}(t_{cmt})$	The time when this review is created
$c_{i,j}(t_{\mathrm{reg}})$	The registration time of the consumer
$c_{i,j}(b)$	The brand of the product

$$\mathrm{CL} = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(c_{i,j}(t_{cmt}) - c_{i,j}\left(t_{pur}\right)\right). \tag{2}$$

3.2.3. Comment Time Interval (CTI). It proves to be a rational expectation that online consumers lack the motivation to compose reviews in a frequent and regular fashion. In contrast, those who are paid do not lack incentives, which contributes to the wide divergence of their frequent review posting behaviors from the common users. A feature targeted at this phenomenon is defined as follows:

$$\mathrm{CTI} = \frac{1}{n_i - 1} \sum_{j=2}^{n_i} \left(c_{i,j}(t_{cmt}) - c_{i,j-1}(t_{cmt})\right), \tag{3}$$

where $c_{i,i}, c_{i,2}, \ldots, c_{i,j}, \ldots, c_{i,n}$ are sorted by t_{cmt}.

3.2.4. Emotional Word and Product Feature Word. Reliable reviews serve to accurately assess the features of products, because actual first-hand information and experience can yield informative and efficient judgments while forging assessments is never easy. In regard to deceptive reviews, a number of optimistic sensible expressions may be present in a bid to flatter some products. How the user chooses words reflects his/her own characteristics. Given that the research sample is a review of cellphones, a professional in this field is consulted to provide a domain-specific word list. Thus, the following three features are built:

POS. Average number of emotionally positive words in consumer reviews

NEG. Average number of negative emotion words in consumer reviews

FEATURE. Average number of product feature words mentioned in consumer reviews

3.2.5. Brand Concentration (BC). The employers of spammers can be a dark company who hires them to commend or tarnish a certain brand; therefore, only a specific type or a small portion of goods are noted with reviews. Therefore, a feature to symbolize this sort of behavior is put forward, where η_k is the percentage of reviews in brand k. This feature is the sum of the Herfindahl index [25].

$$\mathrm{BC} = \sum_{k=1}^{m} \eta_k^2, \tag{4}$$

is each brand's squared market share for the given consumer. If n brands have an equal review share, BC will be $\sum_1^n (1/n)^2 = 1/n$. Provided the entire share is occupied by a single brand of product, BC will be 1. A higher BC feature indicates a more concentrated user, and that user is likely to hold a membership among the Water Army.

3.2.6. Text Length (TL). Paid posters tend to exert a great effort to influence consumers' behavior and imitate normal reviews. However, the common reviewers possess poor motivations to deliver a lengthy and elaborate review. That is why the average review length of a reviewer works well to identify the paid posters among all the reviewers, where $\|c_{i,j}(\text{text})\|$ is the text length of $c_{i,j}$.

$$\mathrm{TL} = \frac{1}{n_i} \sum_{j=1}^{n_i} \left\| c_{i,j}(\text{text}) \right\|. \tag{5}$$

3.2.7. External Text Similarity (ETS). The reviews of the Water Army present the feature of repetition or similarity in content and pattern. It is odd to find two individual users publishing the same exact text, so the similarity among reviewers shall be cited as a feature.

$$\mathrm{ETS} = \frac{1}{n_i} \sum_{j=1}^{n_i} \left\| \left\{ c \mid c \in \mathbf{C}, c(\text{text}) = c_{i,j}(\text{text}) \right\} \right\|. \tag{6}$$

3.3. Dataset. The following two datasets are employed in this research to obtain a thorough and objective performance evaluation:

JD. JingDong (JD.com NASDAQ:JD) stands as one of China's largest B2C online retailer bodies; it has 1500 different product categories and more than 20 million products. This dataset was collected by ourselves. Attention was paid to the preprocessing and data cleaning process.

Amazon. This dataset was bestowed to this study by [8] and engages a visible variety of goods and covers 5.9 million online product reviews, 2.24 million users, and 6.8 million different products.

The authors labeled potentially paid posters by scanning their reviews, together with other meta-information (most of them were uninformative or inconsistent). The reason why we use the word "potential" is in an effort to avoid disputes over this contentious fact. Any certified or sure judgment fails to hold water unless the reviewer or his boss make it public, neither of which are inclined to happen. The discussion over a distinct distinction among real and fake users stretches beyond the technical sphere. The sample size is shown in Table 2.

Table 2: Sample size of two datasets.

Dataset	Reviews	Normal users	Paid posters
JD	32386	485	451
Amazon	260711	642	645

The purchase time of each review appears inaccessible in the Amazon data-bank, which is why the CL feature failed to be counted for the Amazon dataset. Although the brand of every commodity is displayed in the Amazon dataset, a large number of missing values remain, and the BC feature proves to be available in the Amazon dataset.

3.3.1. Distribution of Features. The kernel density estimation of each feature is plotted in Figures 3 and 4.

The distribution of features on the two datasets shows similar results. Normal users tend to write different truthful reviews for different products, which means that a majority of normal users will display low ITS features. CTI has substantial distinguishing power. There is a sharp increase in the left part, which means that the paid posters tend to make reviews more frequently. FEATURE also has much distinguishing power as shown by the high peak in the density curve. Normal users pay more attention to the specific features of products. However, paid posters do not care about the products. They write reviews for the sake of profits.

3.4. Analysis of Feature Influence. Figures 3 and 4 give an intuitive description of each feature's distinguishing power but do not provide more detailed explanations. More empirical studies are needed to test whether each feature has a significant influence on the detection of fake reviews. From an econometric view, it is a discrete choice model, so a logistic regression is the most suitable regression model. The result is shown in Table 3.

The following few conclusions can be drawn from the regression results:

(1) Paid posters have a significantly higher tendency to write reviews similar to previous reviews and reviews from other users

(2) Paid posters tend to write reviews more frequently than normal users

(3) Paid posters often use fewer negative emotional words and more product feature words than the normal users

3.5. Classification Methods. The authors establish a novel platform via ensemble modeling and gather seven classification algorithms to differentiate the Water Army from common customers; the included algorithms are the neural network, decision tree, logistic regression, SVM, random forest, stochastic gradient descending, and k-nearest neighbor, as shown in Figure 5. The aforementioned mechanics are totaled by a collective electing system. At the outset, ensemble learning is designed to upgrade a certain model or decrease the prospect of an unexpected option of

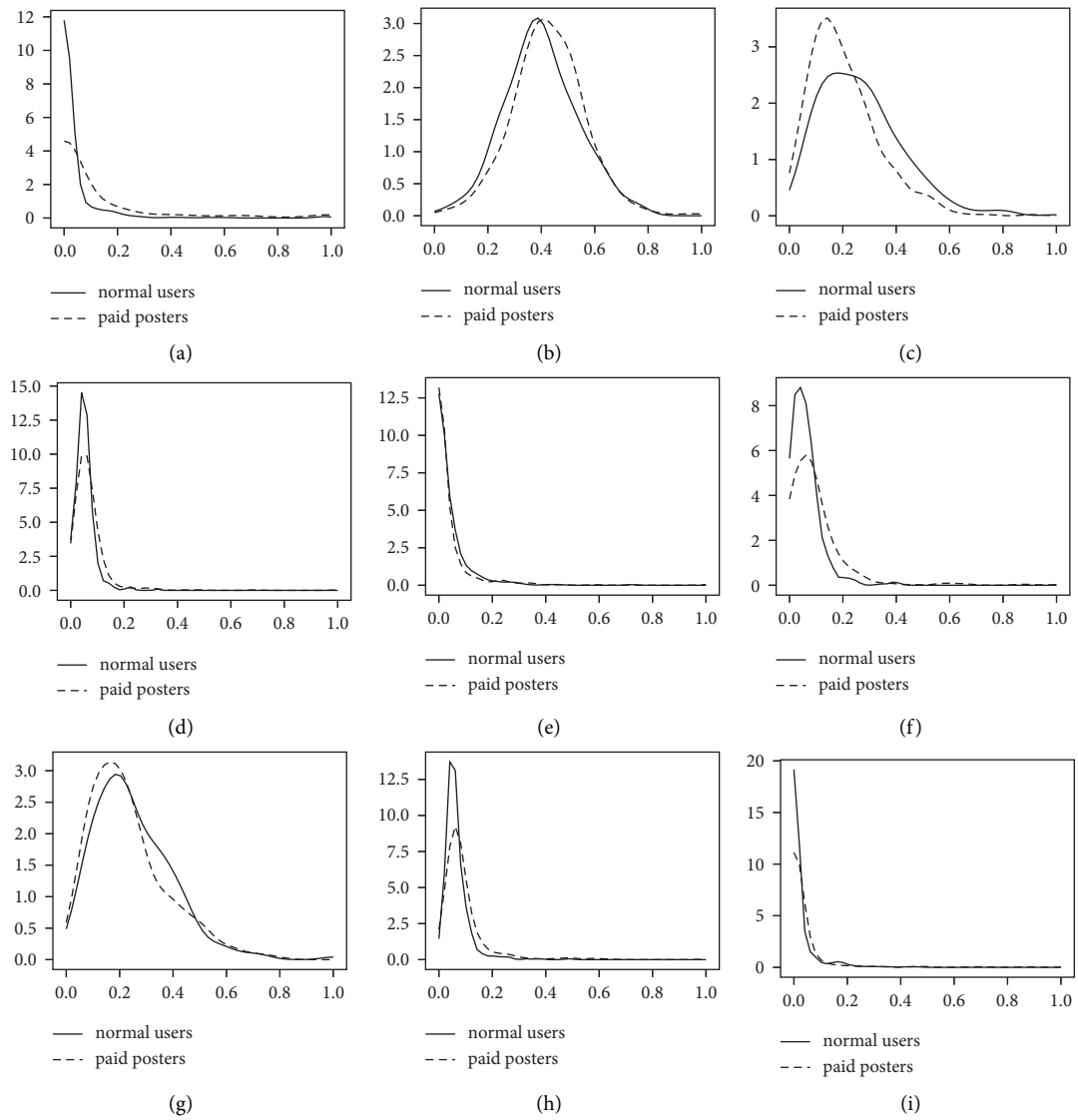

Figure 3: Feature distribution in JD dataset. (a) ITS. (b) CL. (c) CTI. (d) POS. (e) NEG. (f) FEATURE. (g) BC. (h) TL. (i) ETS.

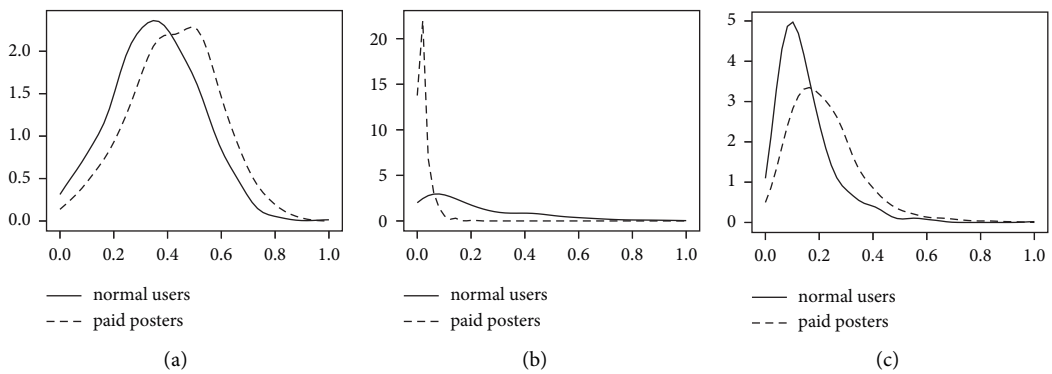

Figure 4: Continued.

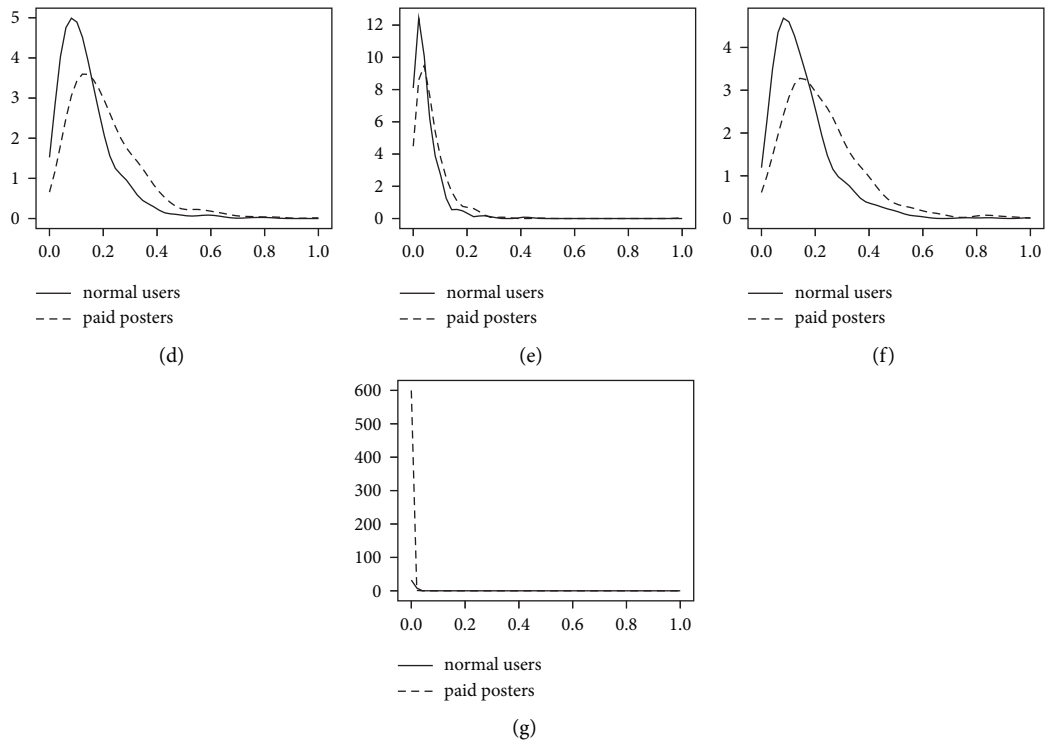

Figure 4: Feature distribution in the amazon dataset. (a) ITS. (b) CTI. (c) POS. (d) NEG. (e) FEATURE. (f) TL. (g) ETS.

Table 3: Logistic regression result.

	JD	Amazon
Intercept	0.789	0.913^{***}
ITS	2.811^{***}	-1.615^{**}
CL	$-1.516e-5$	
CTI	$-7.883e-7^{***}$	$-1.53e-6^{***}$
TL	0.007	0.005
BC	−0.824	
ETS	$2e-4^{**}$	-1.837^{***}
POS	0.118	0.105^{**}
NEG	-2.766^{***}	−0.023
FEATURE	0.456^{***}	−0.382

***, **, and * denote significance at 0.01, 0.05, and 0.1 levels, respectively.

an inefficient classifier. The performance following the implementation of combined classifiers contributes to selection decisions, albeit it is not the finest classifier performance among the ensemble.

4. Evaluation

This section gives a thorough description of the assessment procedure. We evaluate our framework's performance by different metrics; instruments such as the ROC curve are employed to inquire into the distinguishing power.

4.1. Evaluation Metrics. A set of metrics are applied to evaluate the performance of the model, and they are listed as follows:

Precision. Percentage of every positive prediction that is faultless.

Recall. Percentage of every authentic positive observation that is faultless.

F1 Measure. The harmonic mean between precision and recall.

ROC Curve and AUC are performance measurements for classification problems at various threshold settings. While utilizing normalized units, the area under the curve (often plainly cited as AUC) is on par with the likelihood that a classifier will rate an arbitrarily selected positive instance higher than an arbitrarily selected negative instance.

All seven algorithms, as well as the ensemble model, are evaluated by a five-fold cross-validation on the entire

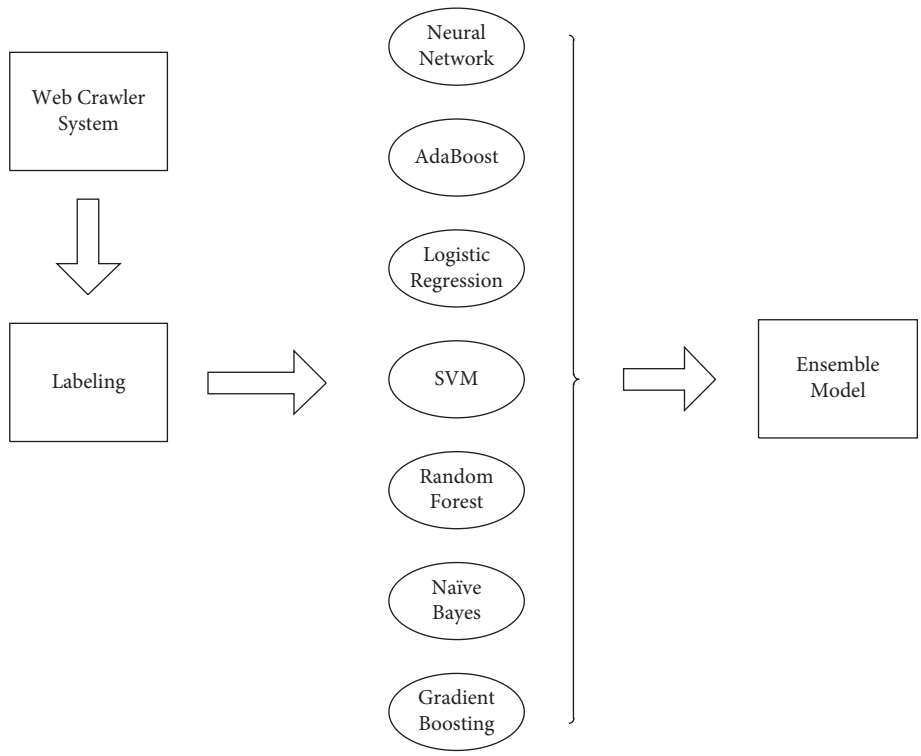

Figure 5: Ensemble model.

dataset. The mean value of all the rounds in the cross-validation process is taken into account as the ultimate count. The outcome of the classification is shown in Tables 4 and 5. On the basis of the results, random forest outperforms the added classification algorithms with a higher $F1$ and AUC scores. Naive Bayes and SVM present a sound performance in the precision measure but score poorly in the recall measure. To rephrase it, many of the paid posters face the prospect of being neglected by the two algorithms. Given the self-acting labeling process feature, a medium magnitude of the sample data proves to be accessible. Consequently, neither algorithmic program demands a large sample database to train the neural network to function at their best in this context. Such a voting mechanism fails to yield visible growth for the ensemble method. Random forest is considered the final classifier. Our detection framework reached 0.730 in the AUC measure, 0.691 in the $F1$ measure on the JD dataset, 0.926 in the AUC measure, and 0.871 in the $F1$ measure on Amazon dataset. To verify whether our algorithm can be applied to large-scale datasets, we also tested the running efficiency of the algorithm. On the JD.com and Amazon e-commerce datasets, the operation efficiency of our ensemble algorithm model is 62.3 seconds and 126.8 seconds, respectively. Because the Amazon dataset contains more commodity data, the running time on this dataset increased.

4.2. Confusion Matrix. Based on the experimental results, we take the random forest as the final classifier of this research and take 20% of the data as the test dataset. The confusion matrix of the experiment is shown in Figure 6.

Table 4: Classification result on JD dataset.

	Accuracy	AUC	$F1$	Precision	Recall
AdaBoost	0.630	0.680	0.629	0.632	0.628
Neural network	0.636	0.703	0.628	0.658	0.583
Gradient boosting	**0.666**	0.720	0.662	0.669	0.646
Logistic regression	0.645	0.698	0.622	0.664	0.588
Naive Bayes	0.589	0.668	0.399	**0.742**	0.274
Random forest	0.651	**0.730**	**0.691**	0.665	**0.659**
SVM	0.606	0.679	0.471	0.722	0.363
Ensemble	0.630	0.704	0.571	0.698	0.455

Table 5: Classification result on Amazon dataset.

	Accuracy	AUC	$F1$	Precision	Recall
AdaBoost	0.861	0.935	0.861	0.839	0.877
Neural network	0.847	0.921	0.849	0.797	0.929
Gradient boosting	0.861	0.924	0.855	0.833	0.895
Logistic regression	0.841	0.901	0.846	0.774	0.949
Naive Bayes	0.811	0.886	0.837	0.738	0.981
Random forest	**0.877**	**0.939**	**0.881**	**0.850**	0.897
SVM	0.771	0.889	0.808	0.687	**0.995**
Ensemble	0.845	0.911	0.865	0.791	0.957

In regard to the general consumers, the detection system operates properly by unmasking the two units. Regarding the Water Army, its precision index decreases, partially as a result of the fluctuation of the annotation process. There is no such thing as a ground truth that can be found or employed as a benchmark to gauge the annotation work. Different people may have dissimilar values on the spamming behavior. What seems to be a reliable review for one

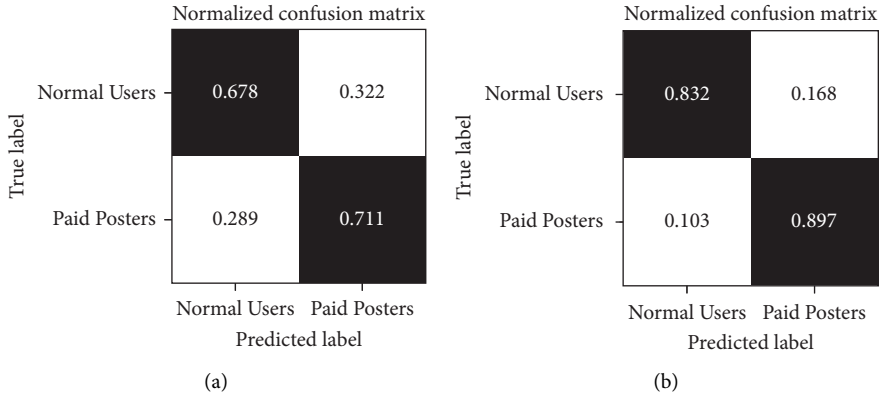

Figure 6: Confusion matrix. (a) JD. (b) Amazon.

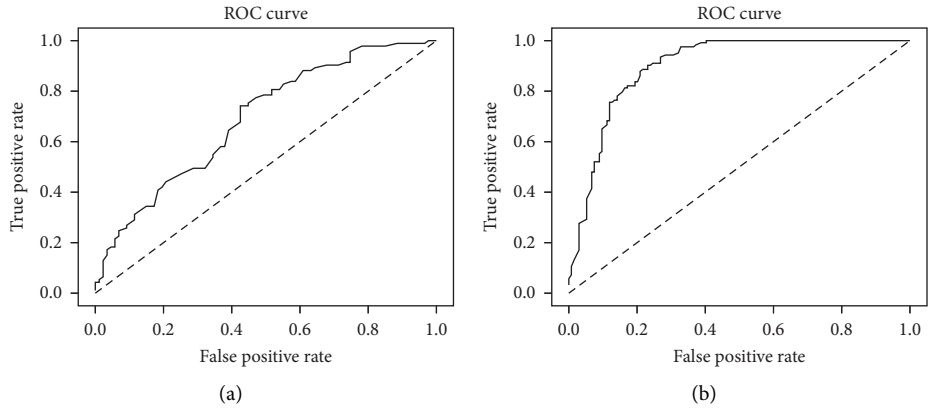

Figure 7: ROC curve. (a) JD. (b) Amazon.

individual may be a fault for another. The inconsistency does weaken our distinguishing power.

4.3. ROC Curve. From the ROC curve, it is seen that the classifier does have a strong distinguishing power. The maximum distance appears against the diagonal line and exists near the heart of the curve. That is why taking 0.5 as the prediction threshold is a sound option, as shown in Figure 7.

4.4. Comparison with Previous Work. The outcomes of this paper are combined with those of the existing studies because of their shared utilization of the Amazon datum. In [8], their AUC measure ranges from 63% to 78%, using different sets of features. In [26], their AUC measure based on a tenfold cross-validation is 78% based on a recent study by Wang et al. From the figures in their paper, the precision fell as the top N sample size grew and exceeded 300. They obtain 95.8% precision in the top 100, 89.6% in the top 200, and 81.8% in the top 300. This detection framework surpasses [27] as the number of observations expands.

5. Conclusion and Future Work

In this paper, the authors probe the identification of the Water Army and advance an extensive set of features to signalize the behavior of the paid posters.

With 4 rating measures on two databanks, the function and operations of this detection framework are thoroughly considered. The AUC and $F1$ of our model reached 0.726 and 0.683, respectively, on the JD dataset, and on the Amazon dataset, the AUC and $F1$ of our model reached 0.926 and 0.871, respectively. Our research has yielded a pragmatic and constructive fix to the paid poster problem from a technological perspective.

5.1. Contribution. Our research makes a strong contribution from both the theoretical and practical perspectives.

First, our research plays a bridging role between the study of e-commerce behavior and the study of fake reviews. Due to the scarcity of fake review datasets, previous scholars have mainly focused on the efficiency and accuracy of supervised and unsupervised learning [11, 27–29]. However, scholars ignore the fact that the game relationship between fake reviewers, ordinary reviewers, and e-commerce platforms may lead to different behavioral strategies. Our study extracted 9 user behavioral characteristics related to fake reviews, which will provide some inspiration for subsequent research on e-commerce fake reviews.

Second, the previous studies are based on single user behavior characteristics, and this study has a relatively comprehensive and novel perspective. We effectively created 9 behavioral characteristics for the false reviewers. From the

results, these features accurately describe the behavioral characteristics of the false reviewers. In addition, considering the differences between domestic and foreign e-commerce platforms, we used the Amazon e-commerce platform data from the United States to verify the robustness of this model. From the results, it is apparent that the model of this study has good results both on the Chinese e-commerce platform dataset and on the American e-commerce platform dataset.

Third, compared with previous studies, scholars generally use a specific classification algorithm to verify the results. Our research innovation is to integrate these verified algorithms. The goal of traditional machine learning algorithms (such as decision tree, artificial neural network, support vector machine, and Naive Bayes) is to find an optimal classifier to separate the training data as much as possible. The basic idea of the ensemble learning algorithm is to combine multiple classifiers to achieve an ensemble classifier with a better prediction effect. The results show that our Ensemble Algorithm achieves good results on both datasets. This provides a novel research idea and method for the study of fake reviews and makes a contribution to the research in this direction.

5.2. Practical Implications. Based on the findings from Sections 3.3 and 3.4, we provide some practical implications that could be applied in the daily operations and product design process of e-commerce websites.

More effort to limit registrations is a sensible decision. The paid posters have a habit of frequently registering a number of accounts as cloaks to mask themselves.

For general users, a real-name registration system should be followed; meanwhile, the cell number and location should be used to take delivery of the reserved products. Whilst for paid posters, they may collaborate with the supplier to lessen the delivery costs by making up a story about the goods An uncompromising real-name system may be implemented and produce effects since general online customers are inclined to accept and follow the related rules. Moreover, the paid posters will face the prospect of a huge economic blow.

Because of the advancement of logistics technology, each and every step leaves a footprint and is recorded in the delivery process via e-commerce web-sites. The logistics information could be combined with the review system as a means to block the paid reviews. A buyer who posts reviews following his confirmation of delivery with reference to the logistics must write the reviews after the logistics notice.

The data in the figures of this study lead us to conclude that the Water Army tends to publish repetitive reviews. Forbidding the copying of reviews may prompt users to exert more effort in writing honest and meaningful reviews.

It is more service friendly to accumulate more suggestive reviews instead of a large amount of invalid or uninformative junk feedback.

5.3. Limitations and Further Research

(1) Our detection frameworks are supervised learning algorithms, which require carefully labeled data as input. The annotation process is labor intensive. Henceforth, the priority of our research lies in reducing the labor cost by adding to the unsupervised learning algorithm channel.

(2) We will proceed to frame a more accurate detection system and then reckon and rate the share of paid posters among all the Chinese shopping websites. Thus, we can provide thorough insights into the behavior of paid posters, as well as offering additional analysis on how much harm the paid posters could do to the industry.

(3) Considering the advancements of fake review research in isolating from data labeling, our followup research will introduce some innovative semi-supervised learning methods that can solve the most challenging and common issues with semisupervised learning, namely, the imbalanced distribution of labeled data over classes [30, 31].

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] China Internet Network Information Center, *The 39th Statistical Report on Internet Development in china*, China Internet Network Information Center (CNNIC), China, 2017.

[2] H. Fang, J. Zhang, Y. Bao, and Q. Zhu, "Towards effective online review sys- tems in the Chinese context: a cross-cultural empirical study," *Electronic Commerce Research and Applications*, vol. 12, no. 3, pp. 208–220, 2013.

[3] S. Utz, P. Kerkhof, and J. van den Bos, "Consumers rule: how consumer reviews influence perceived trustworthiness of online stores," *Electronic Commerce Research and Applications*, vol. 11, no. 1, pp. 49–58, 2012.

[4] N. S. Koh, N. Hu, and E. K. Clemons, "Do online reviews reflect a product's true perceived quality? an investigation of online movie reviews across cultures," *Electronic Commerce Research and Applications*, vol. 9, no. 5, pp. 374–385, 2010.

[5] C. Chen, K. Wu, V. Srinivasan, and X. Zhang, "Battling the internet water army: detection of hidden paid posters," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 116–120, IEEE, Niagara, ON, Canada, August, 2013.

[6] P. P. Chan, C. Yang, D. S. Yeung, and W. W. Ng, "Spam filtering for short messages in adversarial environment," *Neurocomputing*, vol. 155, pp. 167–176, 2015.

[7] K. L. Short, "Buy my vote: online reviews for sale, Vand," *Journal of Entertainment and Technology Law*, vol. 15, p. 441, 2012.

[8] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp. 219–230, ACM, Palo Alto, CA, USA, February, 2008.

[9] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in *Proceedings of the 21st International Conference on World Wide Web*, pp. 191–200, ACM, Lyon, France, April, 2012.

[10] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting burstiness in reviews for review spammer detection," in *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 175–184, Atlanta, GA, USA, June, 2021.

[11] B. Guo, H. Wang, Z. Yu, and Y. Sun, "Detecting the internet water army via comprehensive behavioral features using large-scale e-commerce reviews," in *Proceedings of the 2017 International Conference on Computer, Information and Telecom- munication Systems (CITS)*, pp. 88–92, IEEE, Dalian, China, July, 2017.

[12] N. Jindal, B. Liu, and E.-P. Lim, "Finding unusual review patterns using unex- pected rules," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 1549–1552, ACM, Toronto, ON, Canada, October, 2010.

[13] N. Li and D. D. Wu, "Using text mining and sentiment analysis for online forums hotspot detection and forecast," *Decision Support Systems*, vol. 48, no. 2, pp. 354–368, 2010.

[14] F. T. McAndrew and C. R. De Jonge, "Electronic person perception what do we infer about people from the style of their e-mail messages?" *Social Psychological and Personality Science*, vol. 2, no. 4, pp. 403–407, 2011.

[15] C. Silva and B. Ribeiro, "The importance of stop word removal on recall val- ues in text categorization, in: neural Networks," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 1661–1666, IEEE, Portland, OR, USA, July, 2003.

[16] C. Laorden, I. Santos, B. Sanz, G. Alvarez, and P. G. Bringas, "Word sense disambiguation for spam filtering," *Electronic Commerce Research and Applications*, vol. 11, no. 3, pp. 290–298, 2012.

[17] N. Korfiatis, E. GarcíA-Bariocanal, and S. Sánchez-Alonso, "Evaluating content quality and helpfulness of online product reviews: the interplay of review helpfulness vs. review content," *Electronic Commerce Research and Applications*, vol. 11, no. 3, pp. 205–217, 2012.

[18] Q. Cao, W. Duan, and Q. Gan, "Exploring determinants of voting for the "help- fulness" of online user reviews: a text mining approach," *Decision Support Systems*, vol. 50, no. 2, pp. 511–521, 2011.

[19] F. Wu and B. A. Huberman, "Opinion formation under costly expression," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 1, no. 1, pp. 1–13, 2010.

[20] F. Li, M. Huang, and X. Zhu, "Sentiment analysis with global topics and local dependency," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI*, pp. 1371–1376, Atlanta, GA, USA, January, 2010.

[21] R. Y. K. Lau, S. Y. Liao, R. C.-W. Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detection," *ACM Transactions on Management Information Systems (TMIS)*, vol. 2, no. 4, pp. 1–30, 2011.

[22] F. Li, M. Huang, Y. Yang, and X. Zhu, "Learning to identify review spam," *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, vol. 22, p. 2488, 2011.

[23] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, "Spotting fake reviews via collective positive-unlabeled learning," in *Proceedings of the 2014 IEEE International Conference on Data Mining*, pp. 899–904, IEEE, Shenzhen, China, December, 2014.

[24] S. Y. Bhat, M. Abulaish, and A. A. Mirza, "Spammer classification using ensem- ble methods over structural social network features," in *Proceedings of the 2014 IEEE/WIC/ACM In- ternational Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, pp. 454–458, IEEE, Warsaw, Poland, August, 2014.

[25] A. O. Hirschman, "The paternity of an index," *The American Economic Review*, vol. 54, no. 5, pp. 761-762, 1964.

[26] N. Jindal and B. Liu, "Review spam detection," in *Proceedings of the 16th International Conference on World Wide Web*, pp. 1189-1190, ACM, Perth, Australia, April, 2007.

[27] Z. Wang, T. Hou, D. Song, Z. Li, and T. Kong, "Detecting review spammer groups via bipartite graph projection," *The Computer Journal*, vol. 59, no. 6, pp. 861–874, 2016.

[28] Y. Wu, E. W. Ngai, P. Wu, and C. Wu, "Fake online reviews: literature review, synthesis, and directions for future research," *Decision Support Systems*, vol. 132, Article ID 113280, 2020.

[29] S.-j. Ji, Q. Zhang, J. Li et al., "A burst-based unsupervised method for detecting review spammer groups," *Information Sciences*, vol. 536, pp. 454–469, 2020.

[30] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semisupervised recurrent convolutional attention model for human activity recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1747–1756, 2020.

[31] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 648–660, 2018.