A Method for the Detection of Fake Reviews Based on Temporal Features of Reviews and Comments

—Wenqian Liu
Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

—Jingsha He
Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

—Song Han
Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

—Fangbo Cai
Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

—Zhenning Yang
Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

—Nafei Zhu
Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

(Corresponding author: Nafei Zhu.)

IEEE DOI 10.1109/EMR.2019.2928964

Abstract—*Online reviews and comments after product sales have become very important for making buying and selling decisions. Fake reviews will affect such decisions due to deceptive information, leading to financial losses for the consumers. Identification of fake reviews has thus received a great deal of attention in recent years. However, most websites have only focused on dealing with problematic reviews and comments. Amazon and Yelp would only remove possible fake reviews without questioning the sellers who could continue posting deceptive reviews for business purposes. In this paper, we propose a method for the detection of fake reviews based on review records associated with products. We first analyze the characteristics of review data using a crawled Amazon China dataset, which shows that the patterns of review records for products are similar in normal situations. In the proposed method, we first extract the review records of products to a temporal feature vector and then develop an isolation forest algorithm to detect outlier reviews by focusing on the differences between the patterns of product reviews to identify outlier reviews. We will verify the effectiveness of our method and compare it to some existing temporal outlier detection methods using the crawled Amazon China dataset. We will also study the impact caused by the parameter selection of the review records. Our work provides a new perspective of outlier review detection and our experiment demonstrates the effectiveness of our proposed method.*

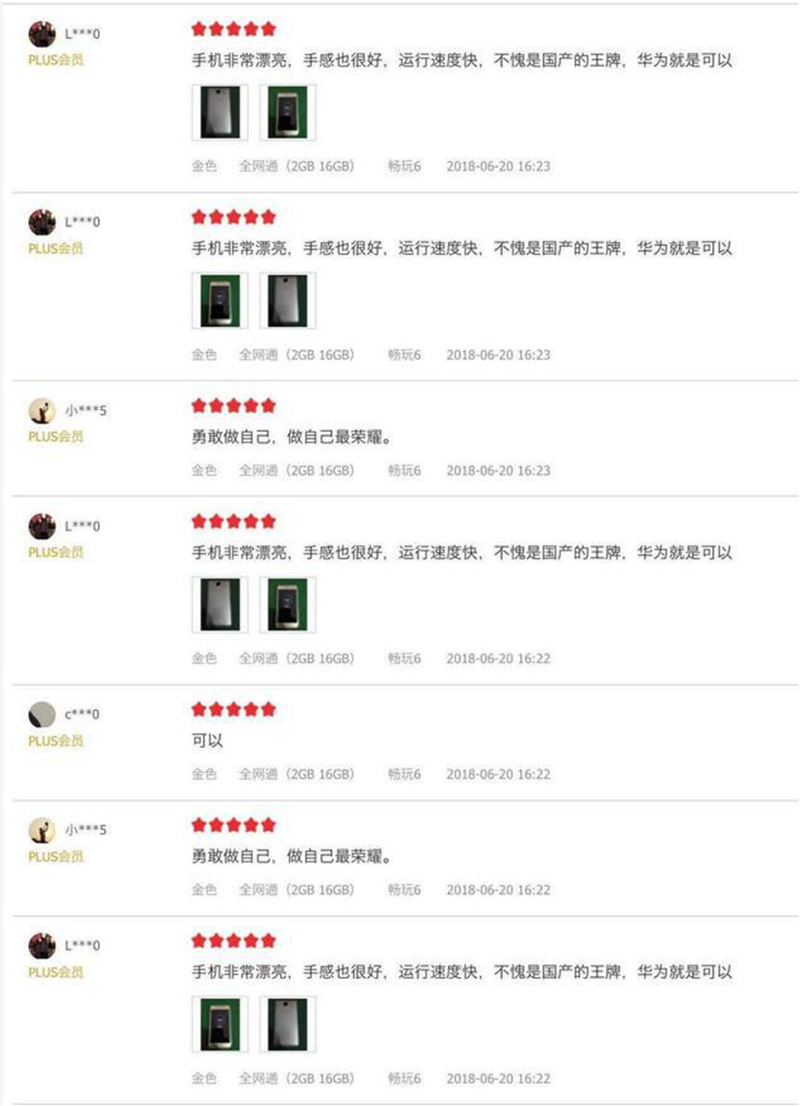
Key words: Fake reviews, products speculation, review records, isolation forest algorithm

1. Introduction

ONLINE shopping has become increasingly popular for individuals and organizations due to the convenience, efficiency and wide merchandise selections that it brings to shoppers. Feedbacks of online shopping have also become an important component for future decision making for both sellers and buyers. Unfortunately, online shopping reviews could be posted without any examination, which may lead to severe problems of deception. For instance, sellers can act as imposters to game the review system and post deceptive fake reviews to promote or upgrade their products and services without much to lose. As shown in Figure 1, a cellphone product has been commented multiple times within a very short period of time by the same user with all the comments being positive with nice pictures. This is very likely a behavior of a colluder. Meanwhile, in our crawled dataset, a product called “净肤面膜” (“Cleaning Mask” in English) enjoyed a rapid growth in the number of reviews within a very short period of time around June 19, 2012, as shown in Figure 2. Such reviews and comments will definitely affect the buying decisions of consumers, giving dishonest sellers the opportunity of earning improper profits.

Even worse, rivals may disguise themselves as buyers to criticize the products and services of competitors. Such individuals are fake customers whose behaviors will prevent shoppers from making civilized judgement. Reviews of such kinds are called fake reviews [1].

In recent years, many methods have been developed for the detection of fake reviews using text mining techniques. Most such work has focused on analyzing one review or one reviewer at a time without considering the potential relationships between multiple reviews or reviewers [2, 3]. Ott et al. employed standard word and part-of-speech features to detect fake reviews from the websites of several different organizations [34]. Han et al. analyzed burst reviews to find the outlier behavior of both reviews and reviewers [4]. Furthermore, the behavior of fake reviewers was analyzed to develop possible review patterns for the purpose of detecting fake reviews [1, 5].



L***0
PLUS会员
★★★★★
手机非常漂亮，手感也很好，运行速度快，不愧是国产的王牌，华为就是可以
金色 全网通（2GB 16GB） 畅玩6 2018-06-20 16:23

L***0
PLUS会员
★★★★★
手机非常漂亮，手感也很好，运行速度快，不愧是国产的王牌，华为就是可以
金色 全网通（2GB 16GB） 畅玩6 2018-06-20 16:23

小***5
PLUS会员
★★★★★
勇敢做自己，做自己最荣耀。
金色 全网通（2GB 16GB） 畅玩6 2018-06-20 16:23

L***0
PLUS会员
★★★★★
手机非常漂亮，手感也很好，运行速度快，不愧是国产的王牌，华为就是可以
金色 全网通（2GB 16GB） 畅玩6 2018-06-20 16:22

c***0
PLUS会员
★★★★★
可以
金色 全网通（2GB 16GB） 畅玩6 2018-06-20 16:22

小***5
PLUS会员
★★★★★
勇敢做自己，做自己最荣耀。
金色 全网通（2GB 16GB） 畅玩6 2018-06-20 16:22

L***0
PLUS会员
★★★★★
手机非常漂亮，手感也很好，运行速度快，不愧是国产的王牌，华为就是可以
金色 全网通（2GB 16GB） 畅玩6 2018-06-20 16:22

Figure 1. Fake reviews for a cellphone.

Most of the methods, however, have relied on analyzing the semantics of shopping reviews. In many situations, complete reviews are not easy to obtain since most of the shopping and reviewing procedures are not really completed. A lot of customers leave the comments to the default or blank, making the comments little useful. Using the record of reviews could provide useful information for the detection of fake reviews. Many recent fake reviews can be featured temporally [4]. In commercial scenarios, fake review behaviors and cheating purchases, where a customer pretends to buy a product and can thus provide reviews and comments, can happen simultaneously which are actually organized by a seller. Such cheating purchase records usually emerge quickly in a short period of time and the behavior can be obtained from the review records.

To capture the review records, we propose an algorithm for fake review detection based on the isolation forest method. In the algorithm, we first analyze the behavior of review records and the outcome of shopping data. We then examine the review records using different window sizes and apply an isolation forest algorithm to deal with the temporal data to capture the outlier reviews. The main idea is to capture or extract the temporal feature that shows the change of patterns over the period of time defined by the size of the window. We will evaluate our method using an Amazon shopping dataset and compare the performance of our outlier detection method to other baseline methods.

Our main contribution lies in the following three folds:

- We empirically analyze the shopping behavior as well as the records of fake reviews.
- We propose an outlier detection method based on temporal features through novel

application of the isolation forest algorithm.

- We evaluate our method using a real world shopping dataset to demonstrate the effectiveness and advantage of our proposed method.

The rest of this paper is structured as follows. In Section 2, we briefly review some related work in feature learning for networks. In Section 3, we empirically analyze the review records and the outlier behavior. In Section 4, we describe the isolation forest based product fake review detection method and in Section 5, we evaluate the proposed method using a real dataset and compare it to several baseline methods to demonstrate the effectiveness and the advantages of our method. Finally, in Section 6, we conclude this paper in which we also point out some promising directions for future research.

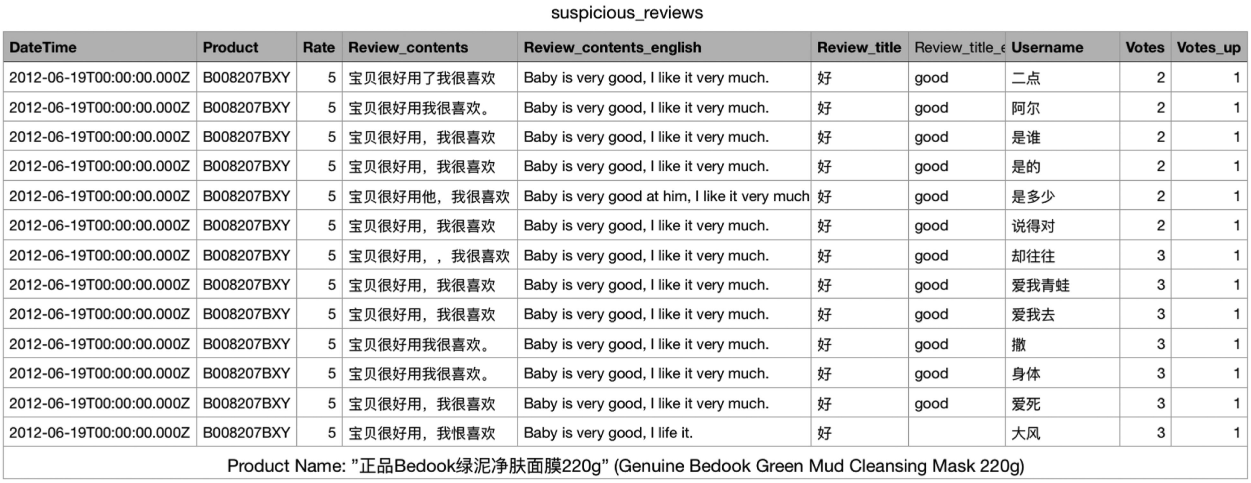
2. Related Work

2.1. Spamming Detection In recent years, Web or email spam have been greatly studied. For example, a survey is provided on Web spam detection [6]. Email spam detection is also studied [7]. Blog spam or network spam are also intensively studied [8, 9]. For the review spams, Fei et al. studied the behavior of fake reviews and provided possible spam patterns [5].

2.1.1. Supervised Spam Detection: Liu et al. were among the first who proposed solutions on the detection of fake or deceptive reviews [33]. In general, fake review detection can be classified into two categories, supervised and unsupervised. Liu provided a detailed survey on spam detection [10]. The method proposed by Liu et al. is a supervised review detection learning method [33] in which fake reviews are regarded as a set of duplicate or near repeat reviews and other reviews are non-fake reviews. A classifier can be built to separate fake and non-fake reviews by employing natural language processing in the detection procedure [3]. The part-of-speech n-gram feature is used to learn the fake and non-fake reviews in a supervised manner. Li et al. proposed to manually label fake and non-fake reviews and then built a classifier over them [11].

2.1.2. Unsupervised Spam Detection:
Unsupervised spam detection was developed to locate the spam without any labeling information. Jindal et al. proposed a rule-based method to mine fake reviews [12]. Lim et al. studied the common patterns of reviews and developed an outlier detection method by using certain predefined review patterns [13]. Wang et al. used a graph-based method to find fake store reviewers by considering the relationships among reviewers, reviews and stores [14]. Xie et al. proposed a method to detect a special group of review spammers who only post one review which is called "singleton" review spam [15]. Such a phenomenon can lead to huge impact on subsequent evaluations, providing a method for the study of the burst review phenomenon for certain users or stores [5]. Moreover, for a group of fake reviewers, each reviewer in the group is more likely to post fake reviews [5]. Some other studies targeted the problem of distributional anomaly of certain reviews to identify the outlier comments by distributional outlier.

2.1.3. Temporal Outlier Detection: Outlier detection, or anomaly detection, has been another thread of study with the goal of identifying patterns in a dataset that don't conform to the regular patterns or are different from the usual behavior. Outlier detection has been developed for a large number of applications in the real world which can be categorized as distance-based, density-based, model-based and isolation-based methods. Different methods have their own advantages in specific



suspicious_reviews

DateTime	Product	Rate	Review_contents	Review_contents_english	Review_title	Review_title_e	Username	Votes	Votes_up
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用了我很喜欢	Baby is very good, I like it very much.	好	good	二点	2	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用我很喜欢。	Baby is very good, I like it very much.	好	good	阿尔	2	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用，我很喜欢	Baby is very good, I like it very much.	好	good	是谁	2	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用，我很喜欢	Baby is very good, I like it very much.	好	good	是的	2	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用他，我很喜欢	Baby is very good at him, I like it very much	好	good	是多少	2	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用，我很喜欢	Baby is very good, I like it very much.	好	good	说得对	2	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用，，我很喜欢	Baby is very good, I like it very much.	好	good	却往往	3	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用，我很喜欢	Baby is very good, I like it very much.	好	good	爱我青蛙	3	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用，我很喜欢	Baby is very good, I like it very much.	好	good	爱我去	3	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用我很喜欢。	Baby is very good, I like it very much.	好	good	撒	3	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用我很喜欢。	Baby is very good, I like it very much.	好	good	身体	3	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用，我很喜欢	Baby is very good, I like it very much.	好	good	爱死	3	1
2012-06-19T00:00:00.000Z	B008207BXY	5	宝贝很好用，我恨喜欢	Baby is very good, I life it.	好		大风	3	1
Product Name: ”正品Bedook绿泥净肤面膜220g” (Genuine Bedook Green Mud Cleansing Mask 220g)									

Figure 2. Fake reviews for product "Cleaning Mask."

applications. With the rapid development of information technology, it has become possible to process a large amount of data through simple outlier detection. However, outlier detection of temporal data deserves more attention. Gupta et al. conducted a comprehensive review of outlier detection methods for temporal data [31] and identified time series for outlier detection as an effective approach. Recently, researchers compared time series data with other similar behavior time series to locate outlier time series in databases [29] or to mine temporal information from social media with time series features [30]. Some methods can be used to detect temporal outliers from data streams [31, 32].

2.1.4. Time Series Outlier Detection: Time series analysis is one of the most actively pursued approaches in outlier detection. Thus, significant amount of work has been carried out in this area. Parametric models for outlier detection in time series first appeared in the late 20th century. Several models proposed in the statistics literature included autoregressive moving average (ARMA), vector auto-regression (VARMA), auto-regressive integrated moving average (ARIMA) and exponentially weighted moving average (EWMA). Direct detection of outliers in time series has been proposed in recent years in which unsupervised discriminative approaches were pursued. These methods rely on the definition of a similarity function that measures the similarity between two sequences and outlier is detected using clustering. By treating all data samples as a time series feature vector, these samples can be clustered and the data sample that is the furthest from all the clusters gets the largest outlier score [16].

Parametric models can also be used to detect outliers in the unsupervised manner where anomalous instances are not specified and a summary model is constructed on the base data. Test data samples can be regarded as outliers if the probability of generating the sequence from the model is unexpectedly low and an anomaly score for the time series can be computed using the model. Popular models of this kind include the hidden Markov models (HMMs) which can be regarded as temporal dependency-oriented mixture models in which hidden states and transitions are utilized to model the temporal dependencies of different time slots. However, HMMs don’t scale well to real life datasets. Also, the training phase of the models heavily depends on the prior selection of model parameters. Furthermore, HMMs are interpretable but cannot scale well to pattern complicated data. Nonetheless, some approaches have been proposed for outlier detection based on HMMs [17–19].

2.1.5. Outlier Detection for Stream Data:
There is another category of methods that perform outlier detection for streaming data where the scenario becomes more complex than regular outlier detection. This is because streaming data can be a time-series or multidimensional and it doesn’t have a fixed length since the incoming data can be infinite. Approaches pursued for traditional time-series are not applicable to streaming data. Methods for outlier detection are similar to conventional high-dimensional outlier detection models, but not with a temporal component.

When dealing with streaming data, evolving prediction models are needed that will update the parameters or model components when new data arrives. These methods can be used to better model the incoming trends of the data. Yamanishi et al. presented a method that employs an online discounting learning approach to learn the probabilistic outlier model in an incrementally manner [20]. The approach includes a decay factor to consider the issue of concept drift. The outlier score for a sample is computed in terms of the probabilistic fit value to the learned mixture model. Such mixture models are used for conventional multi-dimensional data, but cannot be adjusted for temporal decay. It is popular to use clustering to model the data streams. For example, an online clustering method was proposed to detect outlier products [21]. Sometimes, updating the parameters of a model is not enough, leading to a model that can modify itself to incorporate drifts in the data stream. At the same time, an approach was proposed to use dynamic Bayesian networks to model data samples that evolve over time [22]. By adding new state variables, the state of a system can be obtained.

There is another category that uses distance based outliers from data streams at any time instant. Methods in this category perform outlier detection based on sliding windows. As the window for a data stream slides, old data points expire as new data points come in and the number of preceding neighbors of any data point decreases. Angiulli et al. proposed a method to efficiently determine distance outliers using a new data structure named indexed stream buffer (ISB), which can be used for a range of queries [23]. Georgiadis et al. found that maintaining all relationships of data points over time can be expensive [24] and thus exploited an important characteristics of sliding windows to predict the expiration of existing data samples. The problem of distance-based outlier detection can also be solved using dynamic cluster maintenance [25,26].

In this paper, we propose an outlier detection method based on ideas from both stream outlier detection and time series outlier detection. We explore the differences between time series and locate the outlier with isolation forest, an outlier detection method which is usually used in sample outlier detection [33]. Different from previous temporal outlier detection methods, our proposed method models the problem based on isolation forest, which will be more efficient for large corpus of data while achieving better accuracy.

3. Trend Analysis of Product Reviews

In this section, we analyze the shopping review data crawled from Amazon China to illustrate the differences in the reviews and comments of different products.

3.1. General Trend for Product Reviews

In this study, we use the Amazon-China dataset. As can be seen in Figure 3, the number of recorded reviews has been growing rapidly. In 2006, only a few reviews were recorded. As time goes by, more and more reviews and comments were recorded due to two reasons. Firstly, product review has become more popular in recent years. As the result, more customers choose to share their shopping experience online. Meanwhile, as we randomly selected reviewing users and reviewed products, it becomes clear that the more recent the time is, the more active the users are, and it would be easier to correlate users with recent products. So more reviews were posted and recorded in the last year. Specifically, during the last month recorded in the review data, the number of reviews were over 120 thousands, which was ten times the amount of five years ago. The online shopping behavior differs from product to product though.

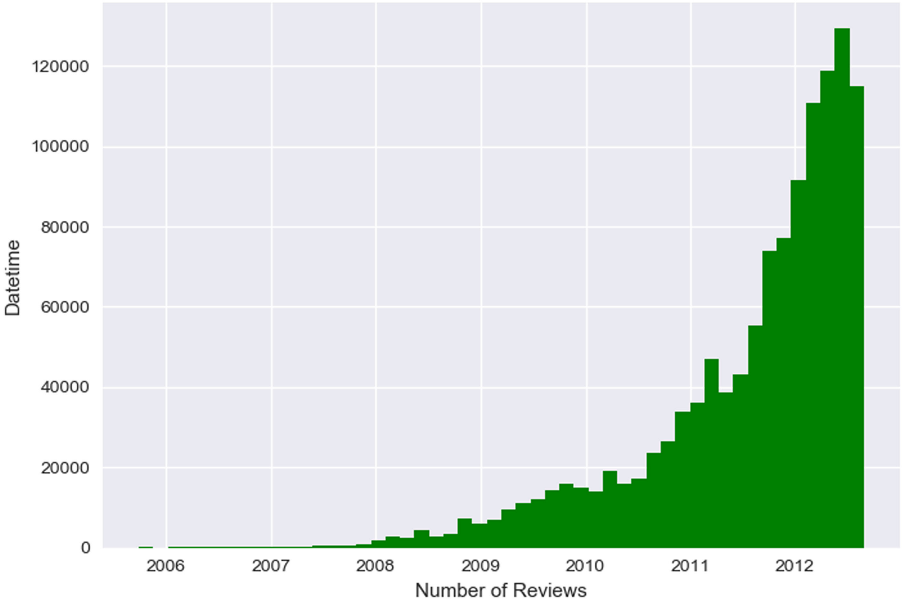


Figure 3. The trend of product reviews.

Table 1. Fake Review Dataset

Information	Value
Products	166,624
User	5,055
Time period	Mar. 2006 - Aug. 2012
Number of reviews	1,205,125
Frequency	507.2 reviews per day

The parameters of the review data is listed in Table 1 in which the dataset contains 166,624 products and 5,055 users and the review period spans between March 2006 and August 2012. In total, there are 1,205,125 reviews, each of which contains a timestamp, a user-product relationship, and a review process. In average, about 500 reviews were recorded each day. The data that we crawled from the Amazon China website can be representative of the general data patterns of shopping websites.

3.2. The Trend in Product Reviews

Figure 4 shows the reviews for four actively commented products selected randomly. First, we define two behaviors: “normal” reviews and “suspect” reviews. Normal reviews are those that are posted by non-colluders in the dataset while suspect reviews are those that are posted by colluders. These four products show different patterns in terms of reviews. Let’s take a look at the temporal patterns of the reviews for the four products. For the products in Figures 4(a) and 4(b), the number of reviews grows very quickly after a short silence. The product in Figure 4(a) is Elizabeth Arden perfume that did not attract too much attention before 2009. After 2009, however, due probably to advertisement or seller initiated promotions, reviews burst and stayed at a high level for the rest of the period. In Figure 2(b), the number of reviews continued to rise due to positive feedbacks by buyers. However, the burst of reviews in both Figures 4(a) and 4(b) is likely caused by sellers who want to maintain high review scores purposely [17, 35]. Meanwhile, as shown in Figures 4(c) and 4(d), normal reviews show a pattern in which the variance of the volumes of reviews between adjacent time intervals is small (generally fewer than 10 reviews in our experiment).

We also selected two other products and analyzed the temporal patterns of the reviews, which is shown in

Figure 5. For the product in Figure 5(a), we can see more clearly that the pattern of trend is not continuous. On the other hand, we plot normal product reviews in Figure 5(b) and can clearly observe that it displays a continuous trend and the number of reviews and comments increases steadily. In Figure 5(c), we summarized the number of suspicious reviews while in Figure 5(d), we plotted the number of normal reviews. There is a clear difference between Figures 5(c) and 5(d) in that the former shows an enormous rise of reviews and comments, indicating a high volume of suspect reviews in history, while the latter shows that the volume of normal reviews and comments rises steadily along with the passage of time.

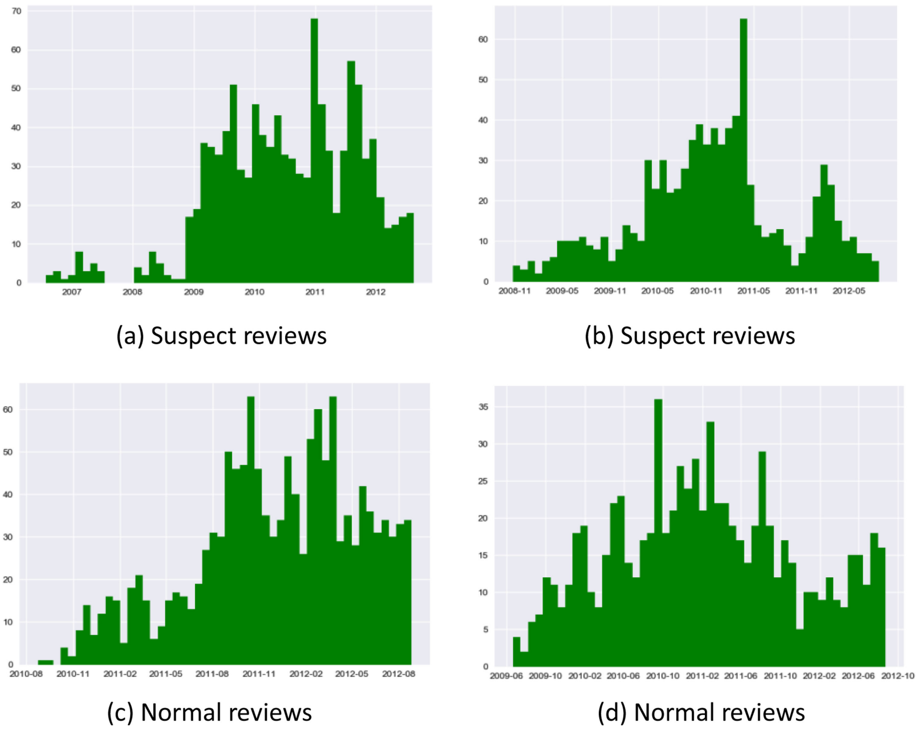
4. The Proposed Method for Detecting Fake Reviews

In this section, we describe our unsupervised fake review detection method. We will also describe how to process the temporal data and use an isolation forest algorithm to discover outlier reviews and comments.

The basic unit of data in our method is a product. We focus on studying the review patterns of all products to obtain outlier products. Since our method for detecting fake reviews is an unsupervised method, fake reviews more likely happen along with other fake reviews and the number of products in one specific time slot could be very large. According to this characteristics, we only collect the reviews within one day as our unit description of one product. For example, for product P_i, the description of the review pattern is defined as:

$$V_{p_i} = [p_{i_1}, p_{i_2}, p_{i_3}, ..., p_{i_E}] \tag{1}$$

where i indicates the date and thus i_1 indicates the first date with reviews in the dataset (March 2006) and i_E indicates the last date with reviews in the dataset (August 2012), and V_{Pi} is the combination feature vector that represents the review records of product P_i. Each element in the vector is a date-review indicator p_{ij} that represents the number of reviews appearing in a certain date i_j. Figure 6 is the general flow diagram of our proposed method for the detection of outlier samples.



(a) Suspect reviews

(b) Suspect reviews

(c) Normal reviews

(d) Normal reviews

Figure 4. Pattern of reviews of four randomly selected popular products.

4.1. Temporal Feature Extraction Based on Reviews and Comments

We view the product reviews statically, and the review records of each product can be processed as an N-dimensional vector. The general form of the temporal feature can be described as:

$$Z_{p_i} = \{z(1)_{P_i}, z(2)_{P_i}, \ldots, z(t)_{P_i}, z(t+1)_{P_i}, \ldots, z(N)_{P_i}\} \tag{2}$$

where$z(t)_{P_i} \in R^N (t \geq 1)$ indicates the number of reviews in time slot t for product P_i and N is the total number of time slots to be processed.

Firstly, all reviews in the same time slot are processed, so are all products in the same time slot. All products will have exactly the same number of time slots. Since the data for the reviews ranges from 2006 to 2012, if we define a time slot as one year, there will be seven time slots in total. It is also important to define all the products with the same dimension. For instance, if the time slot is M days and there are N time slots in total, then $z(t)$ will be:

$$z(t)_{P_i} = \sum_{m=1}^{M} p_{i_{t*M+m'}} \tag{3}$$

where t is the t^{th} time slot of the feature and subscript t^*M+m' indicates the date time of the specific review.

We can then describe all products using a matrix:

$$Z = [Z_{P_1}, Z_{P_2}, \ldots, Z_{P_P}]_{P\times N} \tag{4}$$

where P is the total number of products. Isolation forest algorithm can be used to process the data.

4.2. Isolation Forest Algorithm for Outlier Detection

For the temporal feature vector Z, the initial outlier detection model is developed in which we build the isolation trees in terms of the bootstrap sampling from dataset Z. The ensemble detection model E is composed of L number of iTrees, namely,

$$E = \{E_1, E_2, E_3, \ldots, E_L\} \tag{5}$$

which E_i is built from the data in the ith time slot. The pseudo-code of the iForest algorithm is presented as Algorithm 1.

```
Algorithm 1: iTree (X,t,N)
Input: X - input data, $t$ - number of trees, $N$ - sub-sampling size
Output: a set of $t$ iTrees
Initialize Forest = {};
set iTree height $h = ceiling\ (\log_2 N)$;
for $i = 1$ to $t$ do do
    X' ← sample(X,N);
    Forest ← Forest ∪ iTree(X',0,h);
end
```

In the algorithm, an iForest consists of multiple isolation trees, namely iTree. We know that iTree is created by selecting the temporal review features of a product as well as the value for each feature randomly [27]. At each node in the isolation trees, the instance set is divided into two parts based on the chosen temporal review value. Generally, products with outlier reviews have review records or review values that are very different from the normal products and are easier to be divided than normal products. In the process of isolation, the outlier products are closer to the root and can be more easily divided than the normal products. In order to alleviate the effects imported by the random characteristics in the process of building the isolation forest, the average depth of products in the forest is calculated, which can serve as the anomalous score of the products. The lower the score is, the further away the product is to the normal products, making it a likely candidate as an outlier product. Figure 7 is the flow chart of the algorithm.

In summary, we use the isolation forest algorithm to build the isolation forest based on product review records. Meanwhile, the outlier score can be obtained by applying the isolation forest algorithm.

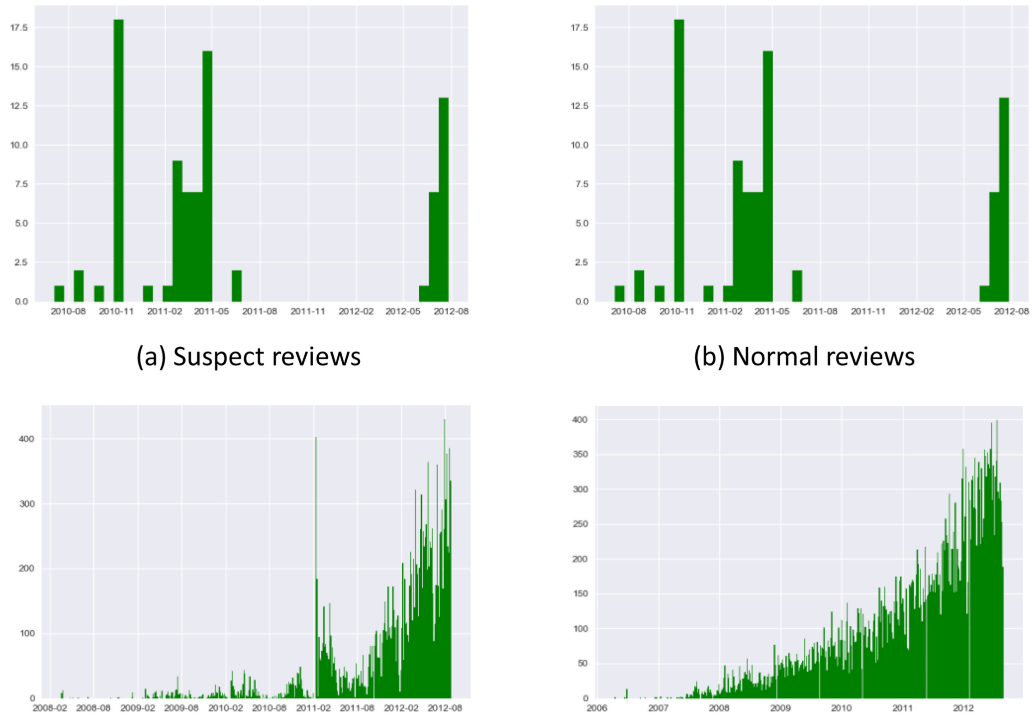
The anomaly score is used to determine whether a product is an outlier product. For product P_i, the anomaly score can be calculated using Formula (6).

$$S(z_{P_i}, N) = 2^{-\frac{E(h(z_{P_i}))}{c(N)}} \tag{6}$$

where

$$E(h(x)) = \frac{1}{L}\sum_{i=1}^{L} h_i(x) \tag{7}$$

In Formula (6), N denotes the sampling size in Algorithm 1, $h_i(x)$ indicates the length of the ith



(a) Suspect reviews

(b) Normal reviews

(c) Suspect reviews

(d) Normal reviews

Figure 5. Pattern of reviews of two typical selected products.

iTree, *E(h(x))* is the average of *h(x)* from a collection of iTrees and *c(N)* is the average of *h(x)* with a given *N*. For product P_i, the outlier score is $S(z_{pi}, N)$. Similar to the results presented in [27], an anomaly score of high value is regarded as an outlier while low value is regarded as a normal sample. A high anomaly score actually indicates a product review pattern that is different from that of normal products temporally. Thus, a high anomaly score is considered to be highly probable that the corresponding product consists of fake reviews and comments. We can therefore use Algorithm 1 together with Formula (6) to analyze the possibility of abnormal reviews of products based on outlier scores. If a product has a high outlier score, it is highly probable that it has fake reviews and comments. On the other hand, a product with a small outlier score can be regarded as being more likely a product without significant fake reviews and comments.

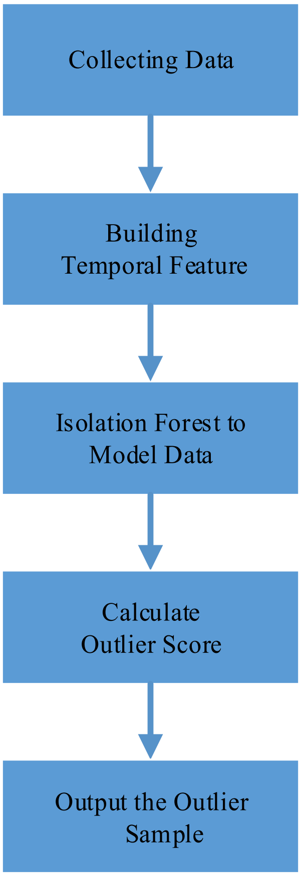

Collecting Data
Building Temporal Feature
Isolation Forest to Model Data
Calculate Outlier Score
Output the Outlier Sample

Figure 6. Flow diagram of the proposed fake review detection method.

5. Experiment and Analysis

We have conducted some experiment using the dataset described in Section 3. Through the experiment, we first compare our proposed method to several baseline fake reviews speculation methods to demonstrate the effectiveness of our method. We then study the performance of our method with different temporal parameter settings. All the experiment was conducted using a Lenovo Desktop with i7-4790 CPU and 8 GB memory size.

5.1. Measurement Metrics To evaluate our method, we quantify the performance in terms of the ground-truth outlier labels and the predicted outlier labels. We use two metrics to measure the performance of our method and to compare it to a few other methods.

- Accuracy: This metric is a straightforward way of measuring the performance of certain predictions. Accuracy reaches the best value at 1 and the worst at 0.
- Running Time: This metric describes how fast an algorithm can finish and yield the results, which we use to measure the efficiency of our method in terms of the number of seconds.

5.2. The Baseline Methods To demonstrate the advantages of our proposed method in detecting fake product reviews, we compare our method to the following three baseline outlier detection methods.

- ARIMA (autoregressive integrated moving average). This method is usually used to detect outliers in time-series data. Here, we use the original ARIMA implementation for time-series data to predict outliers for each product.
- LOF (local outlier factor). This is a commonly used outlier detection method based on the measurement of local deviation of the density of a given sample with respect to its neighbors. We use a commonly used setting of parameters for this method.
- SVM (support vector machine): Although this model is not strictly an outlier detection method, it can

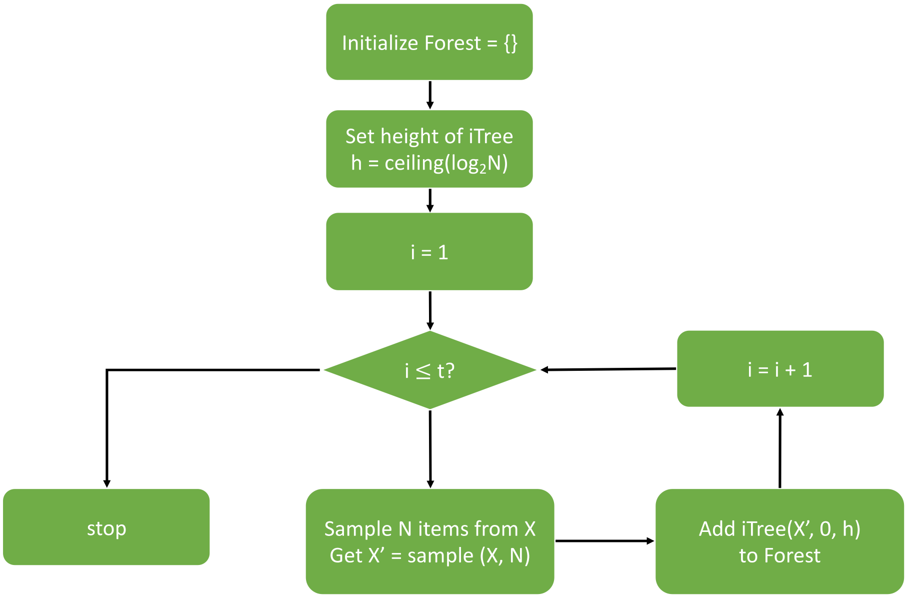


Figure 7. Flow chart of iTree.

still be used to estimate outliers that are not a fit to the general training data. This method is more suitable for complicated and high-dimensional data.

5.3. Comparison Analysis and Evaluation of Experiment Results

In this section, we apply three baseline methods, i.e., ARIMA, LOF and SVM, to detecting outlier products using the dataset described in Section 3. We know that the crawled dataset contains many products that can be detected by the system in [28]. Thus, we follow the same re-crawl strategy to detect whether each product is regarded as normal or abnormal. For each product, a low anomaly score is regarded as abnormal, showing that the product possesses the outlier commercial behavior. The results of accuracy for all the methods are shown in Figure 8, which shows that our method can detect fake reviews with higher accuracy. Lower accuracy means that the detection method is less promising while higher accuracy makes a method a promising candidate to be useful for the detection of fake reviews.

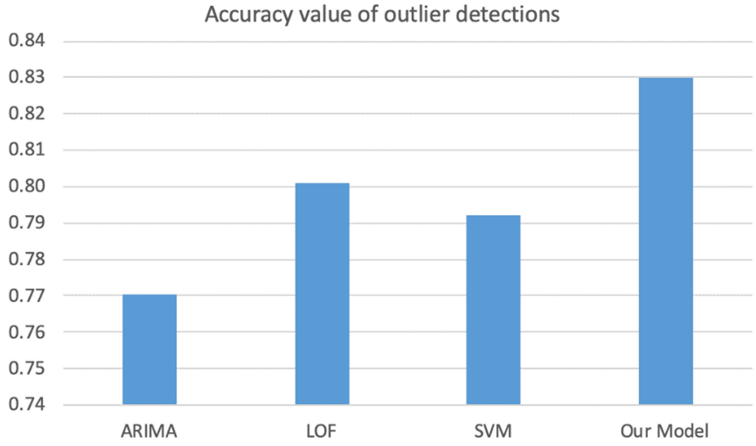


Figure 8. Comparison of accuracy.

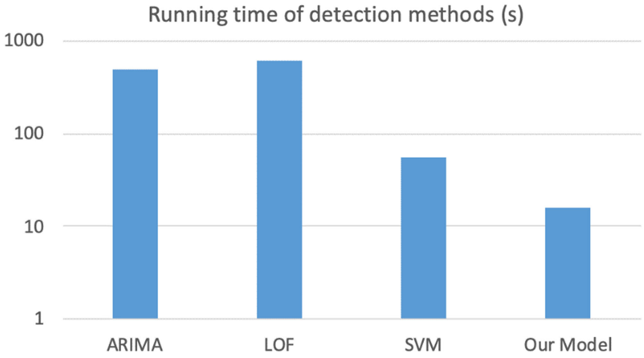


Figure 9. Comparison of efficiency.

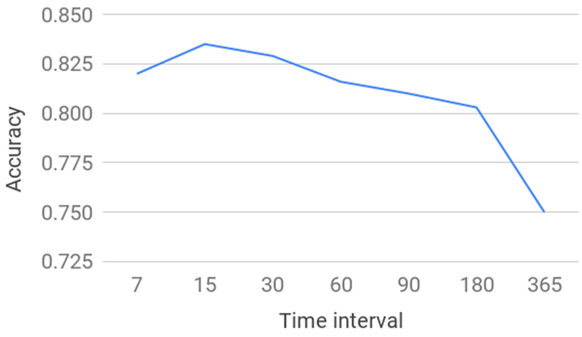


Figure 10. Accuracy vs. time interval.

As can be seen in Figure 8, our method performs better than the other three baseline methods. ARIMA can only reach 0.77 in accuracy, which may be caused by insignificant performance change in time-series. LOF is the most competitive method among the three, indicating that outlier can also happen locally. Our isolation forest-based method demonstrates itself to be superior to all the other three methods. High accuracy means that we can successfully detect the review behavior caused by sellers. The main reason is that our method depends on the neighborhood feature and all product review patterns can be recorded using the iTree, making it possible to detect the variance caused by both popular and unpopular products.

We also compared our detection method to the three baseline methods in terms of efficiency and the results are shown in Figure 9, which shows that our isolation forest-based method can significantly reduce the amount of running time and the approach can be fast in both the training and the evaluation phases.

5.4. Experiment With Varying Time Intervals

We also studied the performance of our method with different time intervals. Experiment results show that the performance of our method is relatively stable for a wide range of time intervals. However, as shown in Figure 10 where the X-axis is the time interval for building the temporal feature vector while the Y-axis is the accuracy, time intervals ranging from 15 to 30 days yield the best performance in terms of the accuracy of detection although the results may not be uniform with the peak value occurring at the time interval of 15 days. This is consistent with some other studies that concluded that most outlier behavior spans from 7 to 15 days [17]. When the time interval is too short, the number of reviews could be vastly different from one time interval to another and outlier detection could be affected by noise in the same interval. When the time interval becomes longer, the difference between adjacent time intervals tends to gradually smooth out, making the outlier behavior less obvious since the volume of reviews may be too large in the same time interval to make noticeable difference.

6. Conclusion

In this paper, we studied the review records of online shopping sites and proposed a novel approach for the detection of fake reviews of products. This review outlier detection method detects the outlier products by analyzing the temporal trends of reviews and comments. Such perspective makes our method more advantageous than some existing methods. Also, detecting review records is efficient and easy. To accomplish the research, we first analyzed the characteristics of online reviews based on an Amazon China dataset. We then proposed an isolation forest based method for fake review detection. We also compared our method with several temporal outlier detection methods to prove the effectiveness and the efficiency of our method.

While our proposed method can better detect outlier products according the abnormal changing trends, the experiment also showed that the selection of window size for the review records plays an important role. A proper window size will yield better performance in terms of the accuracy of detection. We observed that outlier reviews is generally bursting, generally happening in only a few days or weeks. The best detection accuracy can be achieved with a time slot of 15-30 days.

The proposed fake review detection method still needs more study though. As the future work, we will conduct more in-depth study and analysis of review record detection over content based methods and apply fusing review content features into outlier detection to further improve the performance of fake review detection. Furthermore, product recommendation is a promising future work based on honest product reviews and comments.

There are a lot of challenges in the detection of fake reviews based on review records. Our experiment did not indicate clearly when a product has the highest probability of being involved in fake reviews and comments, which is another interesting piece of future work.

Acknowledgment

This work was supported by National Natural Science Foundation of China under Grant 61602456.

References

[1] Streitfeld, D. *Fake Reviews, Real Problem*, The New York Times Company, New York, NY, USA, 2012. http://query.nytimes.com/gst/fullpage.html

[2] Rayana, S.; Akoglu, L. Collective Opinion Spam Detection: Bridging Review Networks and Metadata. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 985–994.

[3] Catal, C.; Guldan, S. Product Review Management Software Based on Multiple Classifiers. *IET Softw.* 2017, 11(3), 89–92.

[4] Zuo, L.; Carass, A; Han, S.; Prince, J.L. Automatic Outlier Detection Using Hidden Markov Model for Cerebellar Lobule Segmentation. *Proceedings of International Conference on Medical Applications in Molecular, Structural, and Functional Imaging*, 2018, pp. 105780D-1-105780D-8.

[5] Mukherjee, A.; Venkataraman, V.; Liu, B.; Glance, N.S. What Yelp Fake Review Filter Might be Doing?, *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, 2013, pp. 409–418.

[6] Spirin, N.; Han, J. Survey on Web Spam Detection: Principles and Algorithms. *ACM SIGKDD Explorations Newslett.*, 2012, 13, 50–64.

[7] Chirita, P.A.; Diederich, J.; Nejdl, W. MailRank: Using Ranking for Spam Detection. *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 2005, pp. 373–380.

[8] Yang, W.; Kwok, L. Improving Blog Spam filters Via Machine Learning. *Int. J. Data Anal. Techn. Strategies*, 2017, 9, 99–121.

[9] Tan, E.; Guo, L.; Chen, S.; Zhang, X.; Zhao, Y. Unik: Unsupervised Social Network Spam Detection. *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, 2013, pp. 479–488.

[10] Liu, B. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Lang. Technol.*, 2012, 5, 1–167.

[11] Li, F.; Huang, M.; Yang, Y.; Zhu, X. Learning to Identify Review Spam. *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, 2011, pp. 2488–2493.

[12] Jindal, N.; Liu, B.; Lim, E.P. Finding Unusual Review Patterns Using Unexpected Rules. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 2010, pp. 1549–1552.

[13] Lim, E.P.; Nguyen, V.A.; Jindal, N.; Liu, B.; Lauw, H.W. Detecting Product Review Spammers Using Rating Behaviors. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 2010, pp. 939–948.

[14] Wang, G.; Xie, S.; Liu, B.; Philip, S.Y. Review Graph Based Online Store Review Spammer Detection. *Proceedings of the 2011 IEEE 11th International Conference on Data Mining*, 2011, pp. 1242–1247.

[15] Xie, S.; Wang, G.; Lin, S.; Yu, P.S. Review Spam Detection Via Temporal Pattern Discovery. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012, pp. 823–831.

[16] Takahashi, T.; Hooi, B.; Faloutsos, C. Autocyclone: Automatic Mining of Cyclic Online Activities With Robust Tensor Factorization. *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 213–221.

[17] Fei, G.; Mukherjee, A.; Liu, B.; Hsu, M.; Castellanos, M.; Ghosh, R. Exploiting Burstiness in Reviews for Review Spammer Detection. *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, 2013, pp. 175–184.

[18] Ovhal, K.B.; Patange, S.S.; Shinde, R.S.; Tarange, V.K.; Kotkar, V.A. Analysis of Anomaly Detection Techniques in Video Surveillance. *Proceedings of the 2017 International Conference on Intelligent Sustainable Systems*, 2017, pp. 596–601.

[19] He, X.; Dai, H.; Ning, P. HMM-Based Malicious User Detection for Robust Collaborative Spectrum Sensing. *IEEE J. Sel. Areas Commun.*, Nov. 2013, 31 (11), 2196–2208.

[20] Yamanishi, K.; Takeuchi, J. A Unifying Framework for Detecting Outliers and Change-Points From Nonstationary Time Series Data. *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 676–681.

[21] Aggarwal, C.C.; Yu, P.S. On clustering Massive Text and Categorical Data Streams. *Knowl. Inf. Syst.*, 2009, 24, 171–196.

[22] Smith, D.V.; Timms, G.P.; de Souza, P.A.; D’Este, C. A Bayesian Framework for The Automated Online Assessment of Sensor Data Quality. *Sensors*, 2012, 12 (1), 9476–9501.

[23] Angiulli, F.; Fassetti, F. Distance-Based Outlier Queries in Data Streams: The Novel Task and Algorithms. *Data Mining Knowl. Discovery*, 2007, 20(2), 290–324.

[24] Georgiadis, D.; Kontaki, M.; Gounaris, A.; Papadopoulos, A.N.; Tsichlas, K.; Manolopoulos, Y. Continuous Outlier Detection in Data Streams: An Extensible Framework and State-of-the-Art Algorithms. *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, 2013, pp. 1061–1064.

[25] Cao, H.; Zhou, Y.; Shou, L.; Chen, G. Attribute Outlier Detection Over Data Streams. *Proceedings of International Conference on Database Systems for Advanced Applications*, 2010, pp. 216–230.

[26] Artinez, E.; Fallon, E.; Fallon, S.; Wang, M. ADAMANT–An Anomaly Detection Algorithm for Maintenance and Network Troubleshooting. *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management*, 2015, pp. 1292–1297.

[27] Liu, F.T.; Kai, M.T.; Zhou, Z.H. Isolation Forest. *Proceedings of the 8th IEEE International Conference on Data Mining*, 2009, pp. 413–422.

[28] Xu, C.; Zhang, J.; Chang, K.; Long, C. Uncovering Collusive Spammers in Chinese Review Websites. *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, 2013, pp. 979–988.

[29] Radhakrishna, V.; Kumar, P.V.; Janaki, V.; Aljawarneh, S. A Similarity Measure for Outlier Detection in Timestamped Temporal Databases. *Proceedings of 2016 International Conference on Engineering & MIS*, 2016, pp. 1–5.

[30] Costa, A.F.; Yamaguchi, Y.; Traina A.J.M.; Traina, C. Jr.; Faloutsos, C. RSC: Mining and Modeling Temporal Activity in Social Media. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 269–278.

[31] Gupta, M.; Gao, J.; Aggarwal, C.C.; Han, J. Outlier Detection for Temporal Data: A Survey. *IEEE Trans. Knowl. Data Eng.*, Sep. 2014, 26(9), 2250–2267.

[32] Salehi, M.; Leckie, C.; Bezdek, J.C.; Vaithianathan, T.; Zhang, X. Fast Memory Efficient Local Outlier Detection in Data Streams. *IEEE Trans. Knowl. Data Eng.*, Dec. 2016, 28(12):3246–3260.

[33] Jindal, N.; Liu, B. Opinion Spam and Analysis. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 2008, pp. 219–230.

[34] Ott, M.; Choi, Y.; Cardie, C.; Hancock, J. T. Finding Deceptive Opinion Spam by any Stretch of the Imagination. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*-, 2011, pp. 309–319.

[35] Kaghazgaran, P.; Caverlee, J.; Alfifi, M. Behavioral Analysis of Review Fraud: Linking Malicious Crowdsourcing to Amazon and Beyond. *Proceedings of the 11th International AAAI Conference on Web and Social Media*, 2017, pp. 560–563.

Wenqian Liu received the B.S. degree from Zhengzhou University, Zhengzhou, China. She is currently working toward the M.S. degree with the Faculty of Information Technology, Beijing University of Technology, Beijing, China. Her research interests include behavioral prediction and data analysis.

Jingsha He received the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 1990. He is currently a Professor with the Faculty of Information Technology, Beijing University of Technology (BJUT), Beijing, China. Prior to joining BJUT in 2003, he worked with IBM, MCI Communications, and Fujitsu Laboratories engaging in R&D of advanced networking technologies and computer security. His research interests include methods and techniques that can improve the security and performance of the Internet. He authored more than 290 papers in the above areas.

Song Han is working toward the M.S. degree with the Faculty of Information Technology, Beijing University of Technology, Beijing, China. His research interests include network security and distributed network technology.

Fangbo Cai received the master’s degree from the Faculty of Information Technology, Beijing University of Technology, Beijing, China, in 2016, where she is currently working toward the Ph.D. degree. Her research interests include network security and distributed network technology.

Zhenning Yang is currently working toward the M.S. degree with the Faculty of Information Technology, Beijing University of Technology, Beijing, China. His research interests include network security and distributed network technology.

Nafei Zhu received the Ph.D. degree in computer science and technology from the Beijing University of Technology, Beijing, China, 2012. She was a Postdoctoral Fellow with the Institute of Software, Chinese Academy of Sciences before joining BJUT in 2017. She is currently an Assistant Professor with the Faculty of Information Technology, Beijing University of Technology, Beijing, China. Her research interests include privacy protection and network security.