SpamTracer: Manual Fake Review Detection for O2O Commercial Platforms by Using Geolocation Features

Ruoyu Deng[1] , Na Ruan[1]*, Ruidong Jin[1], Yu Lu[1], Weijia Jia[1], Chunhua Su[2], Dandan Xu[3]

[1] Department of CSE, Shanghai Jiao Tong University, China
{dengruoyu,naruan,tracyking,luyu97,jiawj}@sjtu.edu.cn
[2] Division of CS, University of Aizu, Japan
suchunhua@gmail.com
[3] China Unicom Research Institute, China
xudd18@chinaunicom.cn

Abstract. Nowadays, O2O commercial platforms are playing a crucial role in our daily purchases. However, some people are trying to manipulate the online market maliciously by opinion spamming, a kind of web fraud behavior like writing fake reviews, due to fame and profits, which will harm online purchasing environment and should be detected and eliminated. Moreover, manual fake reviewers are more deceptive compared with old web spambots. Although several efficient methods were proposed in the fake review detection field, the manual fake reviewers are also evolving rapidly. They imitate to be benign users to control the velocity of review fraud actions, and deceive the detection system. Our investigation presented that geolocation factor is potential and can well reflect the distinctions between fake reviewers and benign users. In this research, we analyzed the geolocations of shops in reviews, found the distinct distribution features of those in fake reviewers and benign users, and proposed a *SpamTracer* model that can identify fake reviewers and benign users by exploiting an improved HMM(Hidden Markov Model). Our experiment demonstrated that SpamTracer could achieve 71% accuracy and 76% recall in the unbalanced dataset, outperforming some excellent classical approaches in the aspect of stability. Furthermore, SpamTracer can help to analyze the regularities of review fraud actions. Those regularities reflect the time and location in which online shops are likely to hire fake reviewers to increase their turnover. We also found that a small group of fake reviewers tend to work with plural shops located in a small business zone.

Keywords: O2O Commercial Platform · Manual Fake Review Detection · Geolocation · Hidden Markov Model

* Corresponding author

1 Introduction

With the explosive growth of electronic commerce and social media, O2O (Online To Offline) commerce has become a heated topic in public. O2O refers to the use of online enticement to drive offline sales, and feedbacks from offline consumption can promote the online dissemination of products [18]. As the feedback part in O2O, reviews of experienced users can provide significant reference values for consumers and help them to make decisions. Opinions in reviews are essential to the evaluation and business volume of a target product in current O2O platforms such as Amazon[4], Booking[5], and Yelp[6]. Positive reviews can bring profits and fame, while negative ones are harmful to products. Due to the pursuit of interest, deceptive reviews and fake reviewers appeared. Moreover, the continuous and rapid evolution of social media makes fake reviewers themselves evolve rapidly and pose a significant challenge to the community [3]. It has been a common practice that shops tend to hire fake reviewers to promote themselves secretly. Those kinds of activities are called opinion spam [9].

Prior researchers have been working on manual fake review detection for several years [21]. At the early stage, methods of opinion spam were elementary and easy to identify. Researchers proposed many approaches based on text analysis [20]. Besides, simple machine learning methods could also be used to classify the suspect reviews by analyzing features of reviews and reviewers [15]. Meanwhile, commercial platforms realized the hazard of opinion spam and built their own filtering systems to find deceptive and inferior quality reviews. Those systems helped purify the disordered review environment, but they also prompted fake reviewers to enrich their poor review contents. Even some skilled fake reviewers were able to deceive the detecting system [24]. As the elapse of time, fake reviewers were becoming more and more cautious and tended to disguise as normal users, and those laggard traditional approaches wouldn't work efficiently anymore. The spotlight on manual fake review detection was gradually shifting from text contents to features and patterns. Some features were proved useful in manual fake review detection like time [10], ranking pattern [5], topics [16] and activity volume [6]. These new approaches did provide several new ideas in opinion spam detection.

We exploit a creative *SpamTracer* method to do manual fake review detection by exploiting the geolocation features. Geolocation is potential in manual fake review detection task. Fake reviewers and benign users may have similar geolocation records. However, fake reviewers don't pay much attention to the position order during review fraud actions. Their strange actions appear to be inconsistent with general behaviors of benign users. The different action concepts between fake reviewers and benign users will cause distinctions in the statistics and the frequency distribution of geolocation features. After computing on a partly labeled reviews and reviewers dataset, we found that both fake review-

[4] www.amazon.com
[5] www.booking.com
[6] www.yelp.com

ers and benign users have double peak distributions regarding the geolocation features. Our method can fit the geolocation features well. Some prior works have discussed the practice of geolocation features in manual fake review detection tasks. Zhang et al. [25] used geolocation features in OSNs (Online Social Networks) to detect fake reviewers, and Gong et al. [7] used LSTM model and check-in information in LBSNs (Location-Based Social Networks) for malicious account detection. Their works enlighten us that location information can reflect some review fraud features.

Apart from detecting fake reviewers, we also discussed the feasibility of discovering the time and location regularities of hiring fake reviewers. There exist some rules in online shop's tendency of hiring fake reviewers regarding time and location. For example, online shops tend to hire fake reviewers in the beginning period to accumulate popularities and obtain a higher rank in searching results, etc. We can draw some conclusions that explain some important regularities based on a large scale dataset expanded by SpamTracer.

In summary, our work makes the following special contributions:

1. We exploit geolocation features to do manual fake review detection in O2O commercial platforms. We extracted the geolocation features of shops, and arrange those from the reviews written by the same person in time order.
2. We built a special SpamTracer model to describe the distribution of geolocation features of fake reviewers and benign users. It's creative that SpamTracer receives geolocation features sequences and gives prediction results.
3. We proposed three significant propositions regarding time and location of review fraud regularities. Our experiment confirmed those propositions and gave reasonable explanations.

The remainder of this paper is organized as follows. In Section 2, we introduce the preliminary works. In Section 3, we present the detailed design and construction of SpamTracer model. The dataset, experiment, and evaluation are demonstrated in Section 4. Finally, we conclude our research in Section 5.

2 Preliminaries

2.1 Terminology

To describe our work precisely, we first introduce some definitions as following.

Definition 1. ***Shop:*** *A shop is an officially registered online shop and holds a unique webpage usually. A shop's webpage contains the detailed description of the shop and a large number of reviews of this particular shop.*

Definition 2. ***User:*** *A user is an officially registered account and holds a personal webpage. A user's webpage contains detailed personal profile and reviews that the user has posted.*

Remark 1. In this paper, we categorize all users into two types: **benign users** and **fake reviewers**. **Benign users** are those who post honest reviews, and **fake reviewers** are those who post fake reviews to promote the target shops.

Definition 3. ***Fake review:*** *Fake reviews are reviews posted by fake reviewers. They post fake reviews without offline experiences. Fake reviews contain fabricated text and imaginary stories, are crafted to mislead normal consumers.*

2.2 Classification Algorithms in Manual Fake Review Detection

Spamming behaviors are categorized into several different types like web spam [21], e-mail spam [2], telecommunication spam [23], and opinion spam [9], etc. Manual fake review detection problem belongs to opinion spam. It can be regarded as a binary classification problem. The critical problem is the selection of approaches and models. According to prior researches, there are several main approaches to detect manual fake reviews.

Texture-based Approaches In 2008, when opinion spamming was firstly proposed by Jindal [9], researchers were focusing on the classification and summarization of opinions by using Natural Language Processing(NLP) approaches and data mining techniques. From 2011, researchers tried to improve the methods of text analysis. Ott et al. [17] built an Support Vector Machine(SVM) classifier using text features including unigrams and bigrams. Shojaee et al. [20] focused on the lexical and syntactic features to identify fake reviews, and Chen et al. [4] proposed a semantic analysis approach that calculates the similarity between two texts by finding their common content words. Traditional texture-based approaches are simple, and they can not reach a high efficiency when manual fake reviewers began to enrich their fake review contents.

Feature-based Approaches From 2014, with the rapid development of machine learning, more and more machine learning algorithms are applied on the fake review detection field. Li et al. [12] proposed a PU-Learning(Positive Unlabeled Learning) model that can improve the performance of Dianping[7]'s filtering system by cooperating with Dianping. Kumar et al. [11] proposed an improved SVM model named DMMH-SVM (Dual-Margin Multi-Class Hypersphere Support Vector Machine) to solve web spamming problem. Chino et al. [6] trained a log-logistic distribution model consisting of time interval and activity volume of one's each review to fit users' behavior, and calculated the dispersion of reviews written by different users to identify those who are isolated from the majority. Li et al. [13] proposed an LHMM(Labeled Hidden Markov Model) combined with time interval features to do fake review detection in a sizeable Dianping dataset and gave an excellent result. Feature-based approach is a powerful weapon in fake review detection, but the features need to continually evolve since the fake reviewers are also evolving themselves simultaneously.

[7] www.dianping.com

Graph-based Approaches From 2016, some researchers chose graph models to find the relations among the products, users, and reviews. A detailed graph model can even capture the deceptive reviewer clusters. Agrawal et al. [1] showed an unsupervised author-reporter model for fake review detection based on Hyper-Induced Topic Search (HITS) algorithm. Hooi et al. [8] proposed a camouflage-resistant algorithm FRAUDAR to detect fake reviews in bipartite graph of users and products they review. Chen et al. [5] proposed a novel approach to identify attackers of collusive promotion groups in the app store by exploiting the unusual ranking changes of apps to identify promoted apps. They measured the pairwise similarity of app's ranking changing patterns to cluster targeted app and finally identified the collusive group members. Zheng et al. [26] proposed an ELSIEDET system to detect elite sybil attacks and sybil campaigns. Feature-based approaches mainly focus on feature selection, while graph-based approaches attach more importance to patterns and links.

2.3 Hidden Markov Model

HMM(Hidden Markov Model) is a classic probabilistic graphical model that uses the graph to represent relations among variables. HMM has two states: observation state and hidden state. Hidden states form a sequence, and every hidden state emits one observation state. In the beginning, HMM has an initial state probability to determine which hidden state will be the first. Every time a new state comes after, hidden states may transform to other states by following a certain transition probability, and the hidden state has a certain emission probability of emitting different kinds of observation states. HMM obeys two significant assumptions. One is that each hidden state only relies on the former one. It guarantees the rationality of transition probability. Another is that each observation state exclusively relies on the corresponding hidden state. It ensures the rationality of emission probability. The two assumptions have been widely acknowledged in practice. In conclusion, an HMM can be represented by three parameters: initial state probability, transition probability, and emission probability under the guarantee of two reasonable assumptions above.

There exist some prior works that apply HMM to manual fake review detection task. Malmgren et al. [14] proposed a basic double-chain HMM and used an efficient inference algorithm to estimate the model parameters from observed data. Washha et al. [22] also proved the qualification of using HMM in manual fake review detection work. Li et al. [13] proposed an LHMM(Labeled Hidden Markov Model) combined with time interval features to detect fake reviews in a sizeable Dianping dataset and gave an excellent result.

3 Manual Fake Review Detection Model

3.1 Symbols and Definitions

Table 1 gives a complete list of the symbols used throughout this chapter.

Table 1. Symbols and Definitions

Symbol	Interpretation
x_i	ith location in review sequence of a reviewer
C	Center point
$Distance(A, B)$	The interval distance between two locations A and B
γ_{x_i}	Distance between x_i and C
$f(x; \mu, \sigma)$	Gaussian distribution function with parameters μ, σ
$N(\mu, \sigma^2)$	Gaussian distribution with parameters μ, σ
L	Label variable
$\lambda = \{\mathbf{A}, \mathbf{B}, \pi\}$	Hidden Markov Model
$\mathbf{A}$	Transition probability of HMM
$\mathbf{B}$	Emission probability of HMM
π	Initial state probability of HMM
X_i	The ith observation state
$X_{1:T}$	The observation states from X_i to X_T
P	Probability
Y_i	The ith hidden state
$Y_{1:T}$	The hidden states from Y_i to Y_T
$a_{j,k}$	The element in the matrix of transition probability $\mathbf{A}$
$b_j()$	The distribution of emission probability

3.2 Structure Overview

In this section, we are going to introduce SpamTracer model used for detecting manual fake reviews. Our detection process is shown in Figure 1. First, extracting the geolocation features from the dataset. Then, arranging the feature sequence in time series. Next, inputing the feature sequence into SpamTracer. Finally, we get prediction results from SpamTracer. SpamTracer makes predictions based on the calculation of possibilities. The prediction results given by SpamTracer are responsible for classifying data samples into fake reviewers or benign users.

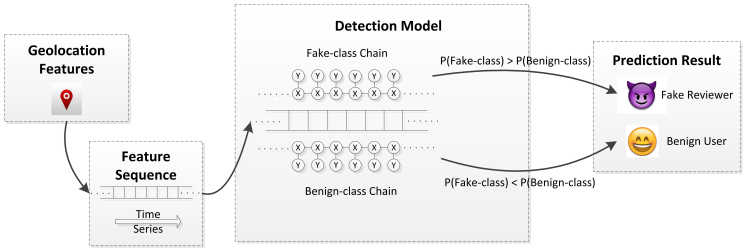


Fig. 1. The structure of manual fake review detection process.

All the symbols and definitions in this chapter are listed in Section 3.1. Then the rationality of the selection of geolocation features will be discussed

in Section 3.3. The methods of modeling geolocation features will be detailedly introduced in Section 3.4. Finally, a discovery of review fraud action regularities will be discussed in Section 3.5.

3.3 Selecting Geolocation Features

Posting reviews is a random process. It means that the posting events are continuously and independently occurring at a constant average rate. Under such process, the related features will follow a particular distribution. As for the feature selection, there were several mature feature distributions discovered by prior work like time intervals and activity volume, etc. However, geolocation features were seldom used in manual fake review detection. The related statistics and the frequency distribution of the location-related feature in fake reviewers and benign users can be calculated and analyzed respectively, and the manual fake review detection problem can be solved by finding the distinctions between them. We use a useful location-related feature, Radius, to measure the disorder degree of users' movement tracks. First, we introduce the definitions of review location, center point and radius:

Definition 4. ***Review location:*** *Review locations are geolocation points of shops that appear in users' reviews. It notes the location where the user purchased offline.*

Definition 5. ***Center point:*** *A center point is the geometric center of the shops in a user' reviews. Determine a user's center contains two steps:*

(1) Find the city that the user lives in by the number of reviews.
(2) Find the geometric center of shops that the user has posted reviews in the city he lives in.

Definition 6. ***Radius****: Radius is the distance between each review location and the center point.*

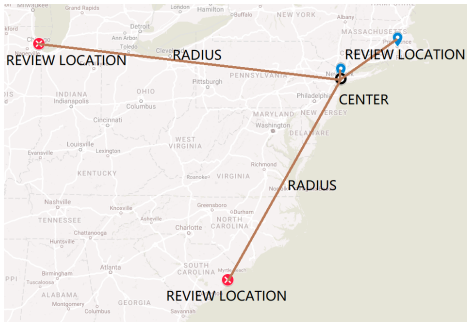


Fig. 2. Definition of radius feature on Google Map.

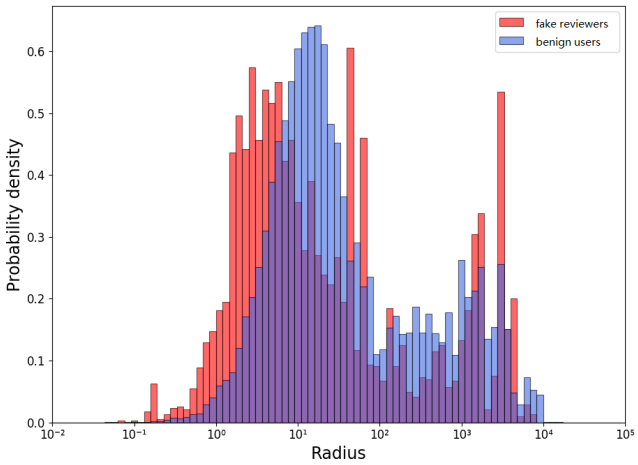


Fig. 3. Frequency distributions of Radius.

Figure 2 shows an example of the definition of radius feature. Most of the review locations are located in New York, so the center point is also located in New York. The lines connecting the center point and each review location represent the interval distances between them, which are the radius for these review locations.

Table 2. Statistics of Radius

	Average Value	Standard Deviation
Fake reviewers	310.0604	678.4959
Benign users	568.5133	999.4281

The statistics and the histograms of the radius calculated on a labeled dataset are shown in Table 2 and Figure 3. The average value and standard deviation show the differences between two reviewer types. The histograms demonstrate that the peaks and slopes are much distinct between fake reviewers and benign users. The frequency distributions can be regarded as the overlap of several Gaussian distributions with different parameters under the log scale x-axis. The double peak distribution pattern is quite reasonable. In general, the range of human activity can be divided into two modes: home range and far range. Benign users tend to purchase near home, and sometimes go far places. It leads to the result that their radius features have the characteristic of double peaks. Although fake reviewers also have two active ranges, they usually take a detour

during review fraud action since fake reviewers don't pay much attention to the location order of fake reviews. The location order of fake reviews written by the same fake reviewer is inconsistent with general behaviors of those belonging to benign users. This is the reason why both fake reviewers and benign users have identical double peak patterns and different peak points and slopes.

The problem can be solved by building a model that can handle with radius sequences. As mentioned above, the distributions of the radius can be seen as the overlap of Gaussian distributions with different parameters. Supposing that $x_i, i = 1, ..., T$ is the location in one's review sequence arranged in time order, the geometrical center C of his most active area can be calculated, then $\gamma_{x_i} = Distance(x_i, C)$ can be used to denote the interval distance between x_i and C, and γ_{x_i} can be drawn from the Gaussian distribution shown in (1).

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad \gamma_x \sim N(\mu, \sigma^2) \tag{1}$$

3.4 Modeling Geolocation features

In this section, we introduce a method of modeling geolocation features and features used to do manual fake review detection work. It's more efficient to deal with data sequences rather than individual data samples because sequences can optimize the differences of action patterns and augment the performance. We proposed a supervised model SpamTracer improved from the classic HMM so that it can deal with the geolocation sequences extracted from the dataset.

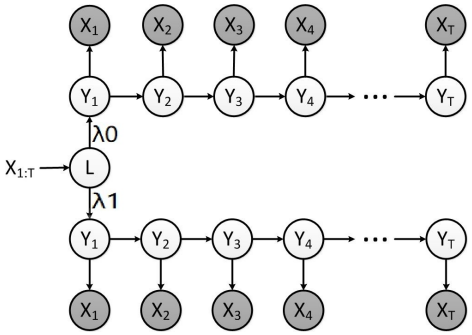


Fig. 4. Representation of SpamTracer.

As illustrated in Figure 4, SpamTracer contains two HMM subchains and a label variable connecting two chains. Label variable is denoted by $L \in \{0, 1\}$, where 0 stands for benign users and 1 stands for fake reviewers. Two subchains $\lambda_0 = \{\mathbf{A}_0, \mathbf{B}_0, \pi_\mathbf{0}\}$ and $\lambda_1 = \{\mathbf{A}_1, \mathbf{B}_1, \pi_\mathbf{1}\}$ represent benign class and fake class, and are trained by two kinds of data samples respectively. When a feature sequence comes, two subchains will calculate the possibility that generates this

sequence. The value of the possibility is a score that measures the fitness between the feature sequence and the class of subchain.

Supposing there is a feature sequence $X_{1:T}$ with unknown label L, the model will calculate its scores under λ_0 and λ_1 respectively, then choose the label by the model that gives a higher score. $X_{1:T}$ also serves as the observation sequences of the subchains. It's rational that the more probable label L takes is the one that generates the observation sequence better.

Our target is comparing the possibility of different label $L \in \{0, 1\}$ under a certain $X_{1:T}$, as expression (2). According to Bayesian theorem, the calculation of $P(L = l|X_{1:T})$ can be converted to the calculation of $P(X_{1:T}|\lambda_l)$ under different λ. First, denominator $P(X_{1:T})$ is independent of L. Thus it is a constant value and won't affect the comparison result, and then it can be dropped. Next, it's easy to get the value of $P(L = l)$ by counting the number of each kind of samples in the dataset. Therefore the problem is the calculation of $P(X_{1:T}|\lambda_l)$. It's equal to the calculation of $P(X_{1:T})$ under different subchains. The detailed calculation process will be introduced next.

$$\begin{aligned} \widehat{L} &= \max_l P(L = l|X_{1:T}) \\ &= \max_l \frac{P(X_{1:T}|\lambda_l) \cdot P(L = l)}{P(X_{1:T})}, l \in \{0, 1\} \end{aligned} \tag{2}$$

Supposing that $x_i, i = 1, ..., T$ is the location in one's review sequence arranged in time order, then γ_{x_i} represents the radius feature of each review x_i. In the subchains of SpamTracer, $X_i = \gamma_{x_i}$ serves as the continuous observation variables, and it follows different Gaussian distributions depending on the hidden state. Considering that a hidden state variable Y_i has two possible values $\{0, 1\}$. Hidden variable Y_i denotes the mode of the point x_i, and the set of $\{0, 1\}$ represents home range mode and far range mode respectively.

The initial state probability π is given as $\pi = \{\pi_j\} = \{P(Y_i = j)\}, j \in \{0, 1\}$. According to the first significant assumption that Y_i only depends on Y_{i-1} and is independent of previous hidden states, the transition probability $\mathbf{A}$ is given as $\mathbf{A} = \{a_{jk}\}$, where $a_{j,k} = P(Y_i = k|Y_{i-1} = j), j, k \in \{0, 1\}$. The observation value X_i is available directly from the dataset. It is emitted by one of the two Gaussian distributions corresponding to the hidden state $Y_i \in \{0, 1\}$. X_i can be demonstrated by (3), where μ and σ are parameters of the Gaussian distribution.

$$X_i = \gamma_{x_i} \sim \begin{cases} N(\mu_0, \sigma_0^2) & Y_i = 0 \\ N(\mu_1, \sigma_1^2) & Y_i = 1 \end{cases} \tag{3}$$

Combined with expression (1), the emission probability $\mathbf{B} = \{b_j(X_i)\}$ can be calculated as expression (4).

$$b_j(X_i) = b_j(\gamma_{x_i}) = P(\gamma_{x_i}|Y_i = j) = f(\gamma_{x_i}; \mu_j, \sigma_j) \quad j \in \{0, 1\} \tag{4}$$

Now the calculation of $\lambda = \{\mathbf{A}, \mathbf{B}, \pi\}$ in the subchains has been stated, as well as how they fit the distribution of radius γ_{x_i}. Supposing $X_{1:T}$ denotes

the observation variable sequence from x_1 to x_T, and $Y_{1:T}$ denotes the hidden variable sequence from x_1 to x_T. Expression (5) formulate the joint probability of $X_{1:T}$ and $Y_{1:T}$. The calculation of $P(X_{1:T}, Y_{1:T})$ means the calculation of the probability that both $X_{1:T}$ and $Y_{1:T}$ appear at the $1 \sim T$ places in order. The time complexity is $O(T)$.

$$\begin{aligned} & P(X_{1:T}, Y_{1:T}) \\ =& P(Y_1, X_1, Y_2, X_2, ..., Y_T, X_T) \\ =& P(Y_1) \prod_{i=1}^{T} P(X_i|Y_i) \prod_{i=2}^{T} P(Y_i|Y_{i-1}) \\ =& \pi_{Y_1} \prod_{i=1}^{T} b_i(X_i) \prod_{i=2}^{T} a_{Y_i Y_{i-1}} \end{aligned} \tag{5}$$

However, the corresponding $Y_{1:T}$ is unknown when given a certain $X_{1:T}$ in SpamTracer. Therefore, all the possible hidden states need to be taken into consideration. SpamTracer needs to calculate 2^T different possibilities of the sequence $Y_{1:T}$. In this situation, the probability $P(X_{1:T})$ can be calculated as expression (6):

$$\begin{aligned} & P(X_{1:T}) \\ =& \sum_{Y_{1:T}} P(X_{1:T}, Y_{1:T}) \\ =& \sum_{Y_{1:T}} P(Y_1) \prod_{i=1}^{T} P(X_i|Y_i) \prod_{i=2}^{T} P(Y_i|Y_{i-1}) \end{aligned} \tag{6}$$

If directly calculating $P(X_{1:T})$ by following the approach above, the time complexity will be $O(T \cdot 2^T)$. Such high complexity is almost uncomputable. In this case, a dynamic algorithm named Forward-backward algorithm [19] was proposed to solve the estimating problem by reducing the time complexity to linear time.

As a result, the calculation of $P(X_{1:T})$ under different subchains is proved practicable. SpamTracer is theoretically qualified as a supervised model and can make predictions. The prediction result given by SpamTracer can be regarded as a score measuring the fitness of data samples and different classes.

3.5 Application of Fake Review Detection Model

In this section, we discuss the review fraud regularities exploration from the dataset with the assistance of SpamTracer. Several empirical conclusions are spreading in public about how to identify fake reviewers. For example, fake reviews hold a large part in the beginning period of most online shops, and there are some periods when fake reviews regularly burst, etc. Besides, fake reviewers tend to look for restaurants competing with others in the same business zone to persuade them to use their review fraud services. As an owner of a restaurant in

a hot business zone, it's easy for him to be forced to hire fake reviewers when he finds that rivals around here are all working with fake reviewers. Fake reviews can be recognized better if their action regularities are revealed. Contraposing to those hypotheses, SpamTracer, and the dataset can tell whether those empirical rules are rumors or truths.

The expansion of labeled data is essential in the review fraud regularities exploration. A more substantial amount of labeled data can lead to much more reliable results. After dataset expansion, all the reviews are labeled, and all the fake reviews are clearly exposed to us. Our research mainly concerns three relations among fake reviewers, time and geolocation:

Date Period and Fake Reviewers We consider the relation between the number of daily fake reviews and the date period, and the regular period of fake review burst. First, SpamTracer identifies the unlabeled reviews in the dataset. Then SpamTracer collects all the fake reviews and their posting time. Fake reviews are categorized by weekdays and months, and a line chart is drawn to explore the regular burst periods.

Shop Opening Days and Fake Reviewers We mainly consider the relation between the number of daily fake reviews and shop opening days, and try to validate the proposition that there are more fake reviewers and reviews in the beginning period of shops. Maybe new shops tend to hire some fake reviewers in the beginning days to help them obtain more population and rise rapidly in the rank. The expansion dataset can show us precisely in which stage restaurants are likely to hire fake reviewers. First, the SpamTracer model identifies the unlabeled reviews in the dataset. Then the fake reviews are assembled by the shops they belong to. Since the dataset contains the shops information, the opening date of shops is available, and the interval days between the shop opening day and review posting day can be calculated. Finally, a histogram chart is drawn to find the distribution of fake reviews posting day.

Shared Fake Reviewers and Interval Distances First, we introduce the definition of shared fake reviewers:

Definition 7. ***Shared fake reviewers:*** *shared fake reviewers are the fake reviewers who simultaneously work with plural shops in a small business zone.*

The existence of shared fake reviewers accelerates the competition in small business zones. Fake reviewers try to force shop owners to use their service by cooperating with their competitors. We plan to count the number of shared fake reviewers and the interval distance between two shops where shared fake reviewers appear. The relation between them can be discovered by drawing a distribution chart.

A simple Algorithm 1 is proposed to calculate the shared fake reviewers of each pair of shops. A review number threshold is set to simplify the computation

Algorithm 1: Calculate the interval distance and amount of shared fake reviewers between two shops

Input: Set of shops M, set of reviewers R, Threshold of shop review number δ;
Output: Pairs of distances and amount of shared fake reviewers $H(d, n)$

1 **for** $\forall m \in M$ **do**
2 **if** $m.numberOfReviews < \delta$ **then continue**;
3 $R_m = \{r \in R | m.reviewer = r\}$;
4 Run SpamTracer on R_m to get the class of every reviewer stored as $r.status$;
5 Add m into set M', M' is a set stores all the useful shops;
6 **end**
7 **for** $i = 1, ..., length(M')$ **do**
8 **for** $j = 1, ..., i$ **do**
9 $d = Distance(M'[i].reviewer, M'[j].reviewer)$;
10 $n = 0$;
11 $List = M'[i].reviewer + M'[j].reviewer$;
12 Sort List by the name of reviewers;
13 **for** $k = 1, ..., length(List) - 1$ **do**
14 **if** *List[k].name = List[k+1].name && List[k].status = fake && List[k].shop != List[k+1].shop* **then** $n = n + 1$;
15 **end**
16 Add (d, n) into set H;
17 **end**
18 **end**

cost. Only those shops hold a certain degree of review numbers can be included in our calculation. First, the algorithm filters shops with fewer reviews, assembling reviewers by shops, and runs the classification model to get their labels(line 1-6). After all the useful shops are prepared, the algorithm travels all the shop pairs and calculates the interval distance and the number of shared fake reviewers(line 7-18). The time complexity is $O(klog(k)n^2)$ where k is the average number of reviews in every shop, and n is the number of useful shops. k can be regarded as a constant value, so the time complexity is $O(n^2)$.

In conclusion, the three propositions are validated with the assistant of dataset expanded by SpamTracer. Some charts demonstrating the links among fake reviewers, time and space will be displayed in the experiment chapter.

4 Experiments

4.1 Dataset Description

Our experiment is based on a Yelp dataset used by Santosh et al. [10]. It is a partly labeled dataset, contains location information, and its reviews are arranged on each reviewer rather than shops. For the labeled part, each review is

labeled as fake or benign by Yelp's filtering system. The dataset information is shown in Table 3. It contains 3,142 labeled users, all of whose reviews are labeled, out of total 16,941 users and 107,264 labeled reviews out of total 760,212 reviews. As for the label reviews, there are 20,267 fake reviews out of 107,624 labeled reviews. A clear boundary is necessary to classify two kinds of users. We referred to Nilizadeh's work [16], calculated the filter rate(i.e., the percentage of filtered reviews out of one's all reviews) of each user, and set a boundary filter rate to cluster two kinds of users. The dataset holds a special characteristic that the filter rate of each user is distributed either in the range of 0-20% or the range of 90%-100%. To separate fake reviewers and benign users, we set a filtering standard that we regard users whose filter rates are higher than 90% as a fake reviewer and lower than 20% as a benign user. Under this standard, there are 1,299 fake reviewers out of 3,124 labeled users. Also, users holding few reviews need to be excluded from the dataset to decrease the unexpected errors. There are 1,796 labeled users and total 11,917 users left if we set the review number threshold as 5.

Table 3. Dataset information

	labeled	total
reviews	107624	760212
users	3142	16941
fake reviews	20267	N/A
fake reviewers	1299	N/A
users after filtering	1796	11917

We rely on the Yelp filtering system for the label work. Yelp filtering system creates the ground-truth dataset and can automatically filter some typical inferior quality and fake reviews. These officially labeled reviews are qualified as the ground-truth dataset. Some prior works used manually labeled data for fake review detection task. However, manual work is not only tedious but also much too subjective. Manual labels are difficult to lead to excellent results.

4.2 Model Evaluation

In this section, the experiment implementation and evaluation of SpamTracer will be presented. The geolocation features are calculated by latitudes and longitudes of every review shop. They are translated from *Arcgis* Map addresses by a python package named *geocoder*. Parameters of SpamTracer are trained from the training dataset, and the evaluation is based on the testing dataset. Training dataset and testing dataset are disjointed parts in labeled data. The ratio of fake reviewers and benign users in labeled data is unbalanced, which is about 1:3. The number of fake reviewers in real O2O platforms is also a minority. And classifiers are required to hold the resistance to the interference from the unbalanced dataset. Many traditional classification algorithms can't perform well

in such situation while SpamTracer can tolerate the impact of large misleading data and recognize the minority fake reviewers exactly.

SpamTracer needs to be compared with other existing excellent approaches to show its advantages in performance. Impartially, some traditional supervised classifiers are selected as the comparison group since SpamTracer is supervised. The comparison group contains four typical classification algorithms: NB(Naive Bayes), AdaBoost Classifier, SVM(Support Vector Machine) and MLP(Multi-layer Perceptron). Those comparison models receive several account characteristics(i.e., friends number, reviews number, etc.) from dataset and output the prediction of fake reviewers or benign user. Besides, our experiment uses a 10-fold CV(Cross Validation) to guarantee the evaluation result. All involved models and their results are presented below:

(1) **SpamTracer**: SpamTracer that receives radius sequences and outputs the prediction.
(2) **NB**: A Naive Bayes Classifier.
(3) **AdaBoost**: An AdaBoost Classifier.
(4) **SVM-rbf**: A Support Vector Machine with Radial Basis Function serving as the kernel function.
(5) **MLP-relu**: A Multi-layer Perceptron with Rectified Linear Unit serving as the activation function.

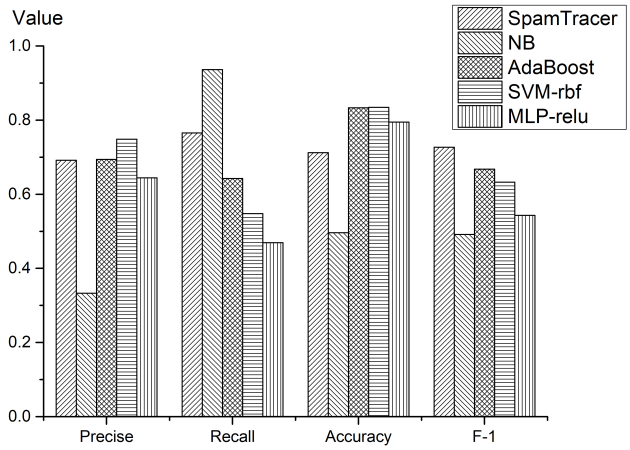


Fig. 5. Precise, Recall, Accuracy, and F1-score of models. SpamTracer performs most stable in all four measures, while other methods fluctuate severely in all four measures.

The evaluation of models is based on four standard performance measures: Accuracy, Precision, Recall, and F1-score. Figure 5 illustrates the four performance measures of all the five models, and shows that SpamTracer performs

Table 4. Precise, Recall, Accuracy, and F1-score data of models

	Precise	Recall	Accuracy	F1-score
SpamTracer	0.6917	0.7657	0.7122	0.7268
NB	0.3332	0.9365	0.4962	0.4916
AdaBoost	0.6945	0.6430	0.8336	0.6677
SVM-rbf	0.7484	0.5482	0.8346	0.6328
MLP-relu	0.6443	0.4694	0.7946	0.5431

most stable in all four measures. NB holds the highest Recall but performs poorest in other three measures. AdaBoost and SVM-rbf perform almost the same as SpamTracer, but they still fluctuate much, and they fall much behind SpamTracer in Recall. MLP-relu holds an excellent Precise and Accuracy while it also holds the worst Recall and F1-score. In summary, SpamTracer is the most stable one in our experiment. Table 4 presents the numerical values of Figure 5.

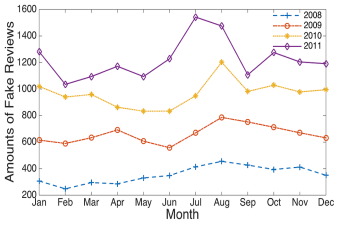
As for the restriction of the performance of SpamTracer, we have some ideas. First, the length of sequences is vital to the performance. The chain structure of SpamTracer determines that the longer data sequences are, the better performance will be. Besides, the scale of dataset also puts a limitation on their performance.

In conclusion, SpamTracer holds excellent stability and performs above the average in all four measures under an unbalanced dataset. Interfered by the unbalanced dataset environment, those classical approaches can't find a compromise among those measures. If we need a stable and precise review filter in O2O platforms, they will not be a good choice since they are likely to miscalculate many normal users or let off many fake reviewers. It's undeniable that SpamTracer will be a better choice for manual fake review detection task.

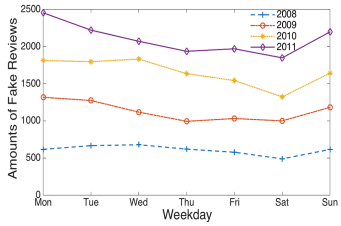
4.3 Regularities of Review Fraud Action

In this section, we are going to state some regularities of review fraud action obtained by applying SpamTracer to the expanded dataset mentioned in Section 3.5. After expanding, there are 694,020 reviews, 228,859 shops and 4,269 fake reviewers out of 11,058 reviewers. All the data samples are labeled. Those reviews mainly covered the period from 2008 to 2011. We mainly concentrate on three relations: fake reviewers and date period, fake reviewers and shop opening days, as well as shared fake reviewers and the interval distance between two shops. Next, we will expound our discoveries and present some figures that can support them.

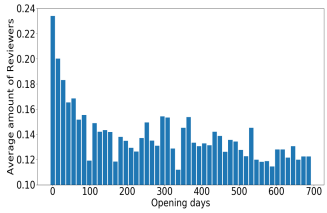
Date Period Fake reviewers tend to burst in summer days of a year and on weekends of a week. The dataset collected the posting date of each fake review and group them by months and weekdays. Figure 6(a) and Figure 6(b) illustrate the month and weekday distribution of fake reviewer bursts from 2008 to 2011.



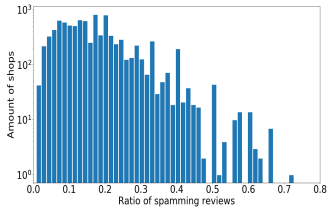
(a) Month and fake reviewers burst



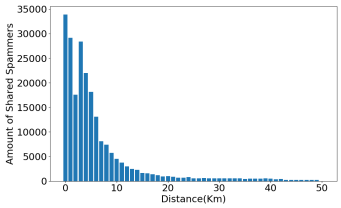
(b) Week and fake reviewers burst



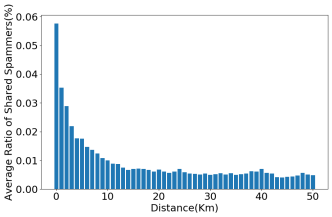
(c) Average fake reviewers and opening days



(d) Shops and ratios of fake reviews



(e) Shared fake reviewers and distance



(f) Average ratio of shared fake reviewers and distance

Fig. 6. Regularities of review fraud action. Figure (a)(b) reflect the date period and fake reviewers, figure (c)(d) reflect the shop opening days and fake reviewers, and figure (e)(f) reflect the interval distance and shared fake reviewers.

Figure 6(a) illustrates that as the elapse of time, fake reviewers tend to make a burst during summer days. According to the data offered by NTTO(National Travel & Tourism Office)[8], most of the overseas tourists visiting the USA came in the 3rd quarter(July, August, and September) from 2008 to 2011. The tourism data reflects a phenomenon that summer is a busy season for traveling. Excessive tourist flows stimulate shop owners to hire more fake reviewers to gain popularity and income. Besides, Figure 6(b) illustrates that fake reviewers tend to write fake reviews on Sunday and Monday. Moreover, both two graphs state a common practice that the amount of fake reviewers is increasing year by year. There were 67,705 fake reviews in 2008 while those grew to 140,722 in 2011. It also reflects that review fraud action is gradually developing in recent years.

[8] https://travel.trade.gov/research/monthly/arrivals/index.asp

Shop Opening Days Fake reviewers most appear in the early stage of shop opening days. Since shops possessing few reviews and shop opening days will interfere final result, we set a filter threshold that only those who have been opening for more than one year and holding more than five reviews are taken into consideration. There are 21,000 shops left after filtering. Figure 6(c) illustrates the final results offered by the dataset and describes the review fraud tendency. X-axis stands for shop opening days, and y-axis stands for the average number of daily fake reviewers in each shop. It shows that more fake reviewers appear in the early shop opening days, and gradually decrease as the elapse of opening days. Besides, we also draw a Figure 6(d) illustrating the number of online shops categorized by the ratio of fake reviews they hold. Figure 6(d) demonstrates that fake reviewers appear in large part of shops. Even there exist some shops whose half of reviews are posted by fake reviewers. It validates a common practice: shops tend to hire fake reviewers to promote themselves secretly.

Interval Distance We discovered a regularity that the amount of shared fake reviewers is inversely proportional to the interval distance between two shops. However, shared fake reviewers only hold a limited percentage of review fraud actions. We set the threshold of shop review number as 2 in our algorithm, and the number of remaining shops after filtering is 102,478. The amount and average ratio of shared fake reviewers are demonstrated in Figure 6(e) and Figure 6(f) respectively. Figure 6(e) shows that there does exist share fake reviewers. However, Figure 6(f) tells that the average ratio of shared fake reviewers is extraordinarily low. It starts from almost 0.06% when the interval distance is nearly 0 and is stabilized at 0.006% with the increase of distance. Two graphs lead to a conclusion that there does exist a phenomenon that some fake reviewers are working with plural shops located in a small business zone, but it's not the main trend of review fraud actions.

5 Conclusion

In this paper, we conducted a research about exploiting geolocation to detect fake reviewers in O2O commercial platforms. We improved a novel detection model, SpamTracer, based on Hidden Markov Model to detect fake reviewers by exploiting the unique distinctions of location features between fake reviewers and benign users. Our evaluation is based on a large scale Yelp dataset and demonstrates that our approach can take manual fake review detection task with excellent accuracy and stability. Also, we discovered some significant regularities in review fraud actions regarding time and location. Fake reviewers tend to launch review fraud actions in the summer season of a year, on weekends of a week, and in the beginning stage of shop opening days. We also found that there existed a negative correlation between the number of shared fake reviewers and the interval distance between two shops.

ACKNOWLEDGMENTS

This work is supported by: Chinese National Research Fund (NSFC) No. 61702330, Chinese National Research Fund (NSFC) Key Project No. 61532013, National China 973 Project No. 2015CB352401, JSPS Kiban(C) JP18K11298 and JSPS Kiban(B) JP18H0324.

References

1. Agrawal, M., Leela Velusamy, R.: Unsupervised spam detection in hyves using salsa. In: Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications. pp. 517–526. New Delhi (2015)
2. Castillo, C., Donato, D., Gionis, A., Murdock, V., Silvestri, F.: Know your neighbors: Web spam detection using the web topology. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 423–430. Amsterdam, The Netherlands (2007)
3. Chakraborty, M., Pal, S., Pramanik, R., Chowdary, C.R.: Recent developments in social spam detection and combating techniques: A survey. In: Information Processing & Management. vol. 52, pp. 1053–1073 (2016)
4. Chen, C., Wu, K., Srinivasan, V., Zhang, X.: Battling the internet water army: Detection of hidden paid posters. In: 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 116–120. Niagara Falls, Canada (2013)
5. Chen, H., He, D., Zhu, S., Yang, J.: Toward detecting collusive ranking manipulation attackers in mobile app markets. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. pp. 58–70. Abu Dhabi, UAE (2017)
6. Chino, D.Y.T., Costa, A.F., Traina, A.J.M., Faloutsos, C.: Voltime: Unsupervised anomaly detection on users' online activity volume. In: Proceedings of the 2017 SIAM International Conference on Data Mining. pp. 108–116. Houston, USA (2017)
7. Gong, Q., Chen, Y., He, X., Zhuang, Z., Wang, T., Huang, H., Wang, X., Fu, X.: Deepscan: Exploiting deep learning for malicious account detection in location-based social networks. In: IEEE Communications Magazine, Feature Topic on Mobile Big Data for Urban Analytics. vol. 56 (2018)
8. Hooi, B., Song, H.A., Beutel, A., Shah, N., Shin, K., Faloutsos, C.: Fraudar: Bounding graph fraud in the face of camouflage. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 895–904. San Francisco, USA (2016)
9. Jindal, N., Liu, B.: Opinion spam and analysis. In: Proceedings of the 2008 International Conference on Web Search and Data Mining. pp. 219–230. New York, USA (2008)
10. KC, S., Mukherjee, A.: On the temporal dynamics of opinion spamming: Case studies on yelp. In: Proceedings of the 25th International Conference on World Wide Web. pp. 369–379. Republic and Canton of Geneva, Switzerland (2016)
11. Kumar, S., Gao, X., Welch, I., Mansoori, M.: A machine learning based web spam filtering approach. In: IEEE 30th International Conference on Advanced Information Networking and Applications. pp. 973–980. Crans-Montana, Switzerland (2016)

12. Li, H., Chen, Z., Liu, B., Wei, X., Shao, J.: Spotting fake reviews via collective positive-unlabeled learning. In: 2014 IEEE International Conference on Data Mining. pp. 899–904. Shenzhen, China (2014)
13. Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., Shao, J.: Bimodal distribution and co-bursting in review spam detection. In: Proceedings of the 26th International Conference on World Wide Web. pp. 1063–1072. Republic and Canton of Geneva, Switzerland (2017)
14. Malmgren, R.D., Hofman, J.M., Amaral, L.A., Watts, D.J.: Characterizing individual communication patterns. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 607–616. KDD '09, New York, USA (2009)
15. Mukherjee, A., Liu, B., Glance, N.: Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st International Conference on World Wide Web. pp. 191–200. Lyon, France (2012)
16. Nilizadeh, S., Labrèche, F., Sedighian, A., Zand, A., Fernandez, J., Kruegel, C., Stringhini, G., Vigna, G.: Poised: Spotting twitter spam off the beaten paths. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. pp. 1159–1174. Dallas, USA (2017)
17. Ott, M., Choi, Y., Cardie, C., Hancock, J.T.: Finding deceptive opinion spam by any stretch of the imagination. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. vol. 1, pp. 309–319. Stroudsburg, USA (2011)
18. Phang, C.W., Tan, C.H., Sutanto, J., Magagna, F., Lu, X.: Leveraging o2o commerce for product promotion: An empirical investigation in mainland china. In: IEEE Transactions on Engineering Management. vol. 61, pp. 623–632 (2014)
19. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. In: Proceedings of the IEEE. vol. 77, pp. 257–286 (1989)
20. Shojaee, S., Murad, M.A.A., Azman, A.B., Sharef, N.M., Nadali, S.: Detecting deceptive reviews using lexical and syntactic features. In: 13th International Conference on Intellient Systems Design and Applications. pp. 53–58. Malaysia (2013)
21. Spirin, N., Han, J.: Survey on web spam detection: Principles and algorithms. In: SIGKDD Explor. Newsl. vol. 13, pp. 50–64 (2012)
22. Washha, M., Qaroush, A., Mezghani, M., Sedes, F.: A topic-based hidden markov model for real-time spam tweets filtering. In: Procedia Computer Science. vol. 112, pp. 833 – 843 (2017)
23. Yao, W., Ruan, N., Yu, F., Jia, W., Zhu, H.: Privacy-preserving fraud detection via cooperative mobile carriers with improved accuracy. In: 2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking. pp. 1–9. San Diego, USA (2017)
24. Yao, Y., Viswanath, B., Cryan, J., Zheng, H., Zhao, B.Y.: Automated crowdturfing attacks and defenses in online review systems. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. pp. 1143–1158. Dallas, USA (2017)
25. Zhang, X., Zheng, H., Li, X., Du, S., Zhu, H.: You are where you have been: Sybil detection via geo-location analysis in osns. In: 2014 IEEE Global Communications Conference. pp. 698–703. Austin, USA (2014)
26. Zheng, H., Xue, M., Lu, H., Hao, S., Zhu, H., Liang, X., Ross, K.W.: Smoke screener or straight shooter: Detecting elite sybil attacks in user-review social networks. In: The 2018 Network and Distributed System Security Symposium. San Diego, USA (2018)