Routledge
Taylor & Francis Group

Journal of Management Information Systems

ISSN: 0742-1222 (Print) 1557-928X (Online) Journal homepage: http://www.tandfonline.com/loi/mmis20

Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning

Naveen Kumar, Deepak Venugopal, Liangfei Qiu & Subodha Kumar

To cite this article: Naveen Kumar, Deepak Venugopal, Liangfei Qiu & Subodha Kumar (2018) Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning, Journal of Management Information Systems, 35:1, 350-380, DOI: 10.1080/07421222.2018.1440758

To link to this article: https://doi.org/10.1080/07421222.2018.1440758

Detecting Review Manipulation on Online Platforms with Hierarchical Supervised Learning

NAVEEN KUMAR, DEEPAK VENUGOPAL, LIANGFEI QIU, AND SUBODHA KUMAR

Naveen Kumar (nkumar7@memphis.edu) is an assistant professor in the Department of Business Information and Technology (formerly Management Information Systems) at the Fogelman College of Business and Economics, University of Memphis. He received his Ph.D. from the University of Washington. His research focuses on deep learning and analytics in social media, information systems, and health care. Before joining academia, he worked as a researcher in the high-tech industry, solving complex problems in information technologies, finance, and manufacturing using machine-learning techniques.

Deepak Venugopal (dvngopal@memphis.edu) is an assistant professor in the Department of Computer Science at the University of Memphis. He received his Ph.D. in computer science from the University of Texas at Dallas. His research interests focus on probabilistic models and statistical relational models. His work has been published in the proceedings of conferences, including those of the Association for the Advancement of Artificial Intelligence, Conference on Neural Information Processing, and others.

Liangfei Qiu (liangfei.qiu@warrington.ufl.edu; corresponding author) is an assistant professor in the Department of Information Systems and Operations Management at the Warrington College of Business, University of Florida. He received his Ph.D. in economics from the University of Texas at Austin. His research focuses on economics of information systems, prediction markets, social media, and telecommunications policy. His work has been published in *Decision Support Systems, Information Systems Research, Journal of Management Information Systems, MIS Quarterly*, and others.

Subodha Kumar (subodha@temple.edu) is the Laura Carnell Chair Professor and director of the Center for Data Analytics at the Fox School of Business, Temple University. He earned his Ph.D. from the University of Texas at Dallas. He has published numerous papers in various journals. He is the deputy editor and a department editor of *Production and Operations Management* and has served as a senior editor of *Decision Sciences*.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/mmis.

Journal of Management Information Systems / 2018, Vol. 35, No. 1, pp. 350–380.

ISSN 0742–1222 (print) / ISSN 1557–928X (online)
DOI: https://doi.org/10.1080/07421222.2018.1440758

ABSTRACT: Opinion spammers exploit consumer trust by posting false or deceptive reviews that may have a negative impact on both consumers and businesses. These dishonest posts are difficult to detect because of complex interactions between several user characteristics, such as review velocity, volume, and variety. We propose a novel hierarchical supervised-learning approach to increase the likelihood of detecting anomalies by analyzing several user features and then characterizing their collective behavior in a unified manner. Specifically, we model user characteristics and interactions among them as univariate and multivariate distributions. We then stack these distributions using several supervised-learning techniques, such as logistic regression, support vector machine, and k-nearest neighbors yielding robust meta-classifiers. We perform a detailed evaluation of methods and then develop empirical insights. This approach is of interest to online business platforms because it can help reduce false reviews and increase consumer confidence in the credibility of their online information. Our study contributes to the literature by incorporating distributional aspects of features in machine-learning techniques, which can improve the performance of fake reviewer detection on digital platforms.

KEY WORDS AND PHRASES: deceptive online reviews, digital platforms, fake reviews, hierarchical supervised-learning, information credibility, machine learning, online reviews, review manipulation.

> *If all you see is five stars, you got to start suspecting these guys work for the manufacturer.*
>
> —Udi Ledergor, CEO, Yotpo [23]

The past decade has witnessed increasing consumer reliance on online reviews [41, 48, 86]. Consider a scenario in which a consumer checks relevant online reviews before buying a new product or trying a new restaurant. One convincing review can often persuade consumers to shift their brand loyalty or drive several extra miles to try a new sandwich shop. However, what if they suspect a particular review has been written disingenuously? They may not follow through with their original purchase or intent. Surveys suggest a majority of modern consumers cite online reviews as a critical factor for planning purchases. Nearly 70 percent of consumers trust the opinions expressed in online reviews, and as many as 90 percent of consumers read online reviews before making financial decisions [54, 65]. Hence, firms have strong incentives to influence their online review ratings: a full-star increase in restaurant online review ratings leads to a 5–9 percent increase in revenue [45], while a half-star increase corresponds to a 19-percentage point reduction in a restaurant's open reservations [5]. For these reasons, both consumers and businesses pay close attention to online reviews about specific products or services on social media.

Two commonly used strategies for influencing online review audiences are management engagement (legal) and deception (illegal). In a formal strategy of online management engagement, business owners may increase the future average ratings of their establishment by offering responses to complaints from low-satisfaction customers [27, 34]. Alternatively, a strategy of online review manipulation

(deception) consists of moves in which business owners inject their public ratings with a positive bias by using fake accounts or paid reviewers [46, 51, 52] or denounce competitors' products [30, 69]. Based on a 2013 poll of social media users, researchers found that 49 percent of respondents acknowledged suspicions that favorable online reviews could be inflated by company-backed incentives such as coupons or credit [6]. In 2015, the research organization Mintel surveyed a sample of consumers and found that 57 percent of consumers are suspicious of companies or products that only have positive online reviews [54]. To this end, we observe that review manipulation causes the deterioration of information quality as well as loss in the credibility of online platforms. In this study, we focus on the detection and prediction of fake reviewers operating on online digital platforms.[1]

Motivation

The phenomenon of online review manipulation is neither new nor exclusive. It is present across multiple industries such as movie, restaurants, hotel, and e-commerce. Hotels or movie studios may use fake accounts or hire real individuals to post overly positive reviews about their movies and attract customers [38, 52]. Some insiders claim that the practice of review manipulation is widely known in the Chinese movie industry.[2] Many movies resort to review manipulation, which can cost more than ¥1 million (around $160,000). Such paid posting is a well-managed activity by Internet public relations companies involving thousands of individuals and tens of thousands of different online IDs.[3] Fake reviews have resulted in lawsuits, costly cases, and eventual settlements [64, 72]. Recently, Amazon has initiated three lawsuits targeting companies that sell positive reviews to vendors [21]. In 2013, New York regulators fined 19 different companies a total of $350,000 for posting fake reviews online [24]. Similarly, Sony was fined by the Connecticut attorney general for creating fake reviews for at least four of its movies [85].

Distinguishing opinion spammers from genuine users in online forums is a challenging, ongoing problem because opinion spammers tend to outsmart genuine users by mimicking their behavior [52]. Cognizant of typical user behaviors within the target review context, opinion spammers generate metadata that resembles that of genuine users. Typically, such fabricated reviews are less easily identified than spam e-mail messages. For instance, e-mails containing variations of the word "Viagra" are almost always spam as most recipients will not find the subject matter relevant.[4] Using imitative techniques, opinion spammers may exercise relatively more influence over users than e-mail spammers. Suppose a user receives an unsolicited email from an unknown person regarding a restaurant. This message will likely be discarded as spam. Online reviews, however, typically do not appear out of context. A user who reads a negative review of a restaurant is likely to trust its message, though written by a stranger. In contrast, unknown e-mail spammers are not met with the same level of trust [75]. Thus, effectively detecting opinion spammers on digital platforms is a challenging problem. Note that detecting opinion spammers (people

who post fake reviews) and detecting spams (fake reviews) are two related topics, but they are not exactly the same. There are subtle differences between the approaches used to detect spam and those used to detect spammers. When detecting spam, the review text is typically one of the most important features that can be used for detection. Several spam detection methods tend to use natural language processing techniques to infer fakeness in terms of writing style [19, 60]. In contrast, for opinion spammer detection, methods that leverage metadata and the joint effect of all reviews written by a user can identify spammers more effectively.

Contribution and Implications

Given the widespread prevalence of online review manipulation phenomena, effective strategies for the positive identification of opinion spammers and deceptive reviews have become critical for firms participating in online review communities. Deception detection is an active area of research and has attracted considerable attention recently in the information systems (IS) domain. The existing studies have mainly focused on predicting and detecting opinion spammers based on supervised-learning techniques with raw or derived features [8, 28, 37, 47, 66, 83]. A few studies employ unsupervised-learning techniques [42, 43, 80], and other nonmachine-learning techniques [36, 61, 74] with raw or derived features as well. Feature engineering plays a critical role in deception detection and prediction. It is the process of transforming raw data into features that better represent the underlying phenomenon under investigation. To the best of our knowledge, the role of feature engineering, more specifically, the distributional characteristics of fake reviewer features, has not been fully exploited in deception detection and prediction.

Opinion spammers tend to distort the underlying natural distribution of features [36]. Only a few studies have analyzed the underlying distributions (e.g., power law, J-shaped distributions, etc.) in online product reviews [14, 32]. Using TripAdvisor hotel reviews and Amazon product reviews, Feng et al. [19] provide insights into the characteristics of natural distributions of opinions. Dalvi et al. [14] analyze the average rating distribution across multiple domains, such as restaurants, movies, and products. They find that the rating distribution is heavily skewed. However, these studies have not specifically focused on understanding underlying distributions from opinion spammers' detection and prediction perspective. In summary, the distributions of features (specific aspects of feature engineering) have not been comprehensively examined to identify potential signals that might detect and predict opinion spammers using supervised-learning methods. This motivates us to incorporate the distributional aspects of reviewers' behavior and interactions in deception detection and prediction.

In this study, we propose a novel approach for spammer detection based on a new feature engineering approach.[5] We develop a hierarchical supervised-learning approach to detect potential opinion spammers, incorporating various univariate and multivariate user-reviewing distribution-based transformation of features

collectively. We build a more robust classifier based on supervised learning, which takes into account the overall review behavior of a user. A striking pattern in online review data is that the features describing opinion spammers are remarkably skewed in nature. This implies that the opinion spammers tend to distort the underlying natural distribution of opinions. Applying our methodology to a real-world data set on online restaurant reviews from Yelp, we demonstrate empirically that it is important to model underlying distributional aspects of reviewer behavior.

Our hypothesis is that features derived from skewed distributions and directly used in the machine-learning algorithms tend to reduce prediction accuracy. We demonstrate this in our baseline model where we do not exploit the natural distribution of features. This is particularly important when we want to discover patterns in data that are skewed, which is the case for most of the review features we consider for our task. We use machine-learning methods including logistic regression, support vector machine (SVM), AdaBoosting, and k-nearest neighbors (k-NN) with the proposed feature engineering approach. We show that using features with the proposed approach increases the likelihood of detecting spammers and also provides comprehensive understanding of a spammer's behavior in the context of online restaurant reviews. The framework of transformation using distributional characteristics of relevant univariate and multivariate features is shown in Figure 1.

In our approach, we first derive several user features related to reviewing behavior, and fit parametric distributions from univariate distribution families that best explain the empirical distribution for those features. Next, we consider the joint distribution of a user's rating and we model this as a Dirichlet distribution. To leverage these distributions collectively within spammer classifications, we transform the space of original feature values to a new feature space that consists of probability density values underlying the features. Unlike traditional classification methods that use raw feature values directly (or scale them using simple methods, such as linear scaling or standardization), our approach has better generalization since the probability density

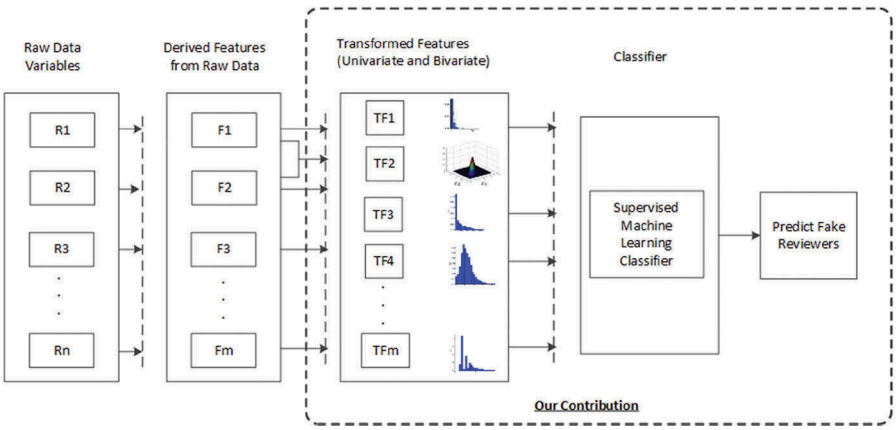

Figure 1. Research Framework

values give the relative importance of a feature value with respect to its true underlying distribution. Our methodology stacks the heterogeneous classifiers together to improve generalization, that is, each distribution by itself is a classifier where the tails of the distribution indicate spammer behavior, and we combine these distributions using a separate meta-classifier.

In addition to our main data set on restaurant reviews, we also validate the robustness of our approach using another data set from a different domain. We collect an additional data set of hospitals from Yelp with the full history of review information for each post, including review date and review rating, for each hospital. We preprocess all the data by applying natural transformation and then learning a supervised model in the new transformed space. Our approach outperforms those that do not use transformed features primarily for two reasons: (1) it enables us to learn models that are nonlinear in nature, and (2) it allows us to understand the univariate and joint behavior clearly.

Our approach has direct implications for digital platforms that are vulnerable to opinion spamming. Our spam detection methodology can be deployed by the platforms to identify spammers and post their information in real time making the technique very effective. Overall, the proposed spammer-detection algorithms supported through digital platforms will help improve consumer experience, ultimately opening more revenue-generating opportunities for both digital platforms and businesses.

Literature Review

We review the literature from three different aspects: (1) firms' engagement in online review deception, (2) deception detection using machine-learning (supervised and unsupervised) methods and nonmachine-learning methods, and (3) feature engineering in deception detection. We also demonstrate our contribution by comparing and contrasting our work with past studies.

Firms' Engagement in Online Review Deception

Fraudulent review manipulation on digital platforms is a serious problem [29]. Firms consistently manipulate online consumer reviews to either promote their products or denounce competitors' products [30]. Several studies have shown the existence of widespread manipulation strategies in online digital platforms. Luca and Zervas [46] find that the positive review manipulation is related to reputational concerns, but they also find that the negative review manipulation is more likely because of competition. Mayzlin et al. [52] explore how hotel characteristics and ownership structure affect the level of review manipulation on travel websites expedia.com and tripadvisor.com using a difference-in-differences approach. Using a game theoretic model, Mayzlin [51] shows that low-quality firms tend to engage in a higher level of review manipulation because high-quality firms benefit from positive word of

mouth, which substitutes for promotional reviews. Hu et al. [29] find that the review manipulation level decreases with the passage of time.

Review manipulation is also seen in strategic recommender systems, and it intentionally affects consumer choices [2]. Forman et al. [20] show that the reviewer disclosure of identity information plays an important role in how community members interpret online reviews. In summary, firms have strong incentives to induce deception by making marketer-generated content and user-generated content difficult to distinguish from each other. However, there is a lack of unified theory and methods for online deception detection [71], motivating us to study and develop more sophisticated methods.

Deception Detection Techniques

The literature related to deception detection techniques can be broadly classified into two research streams: machine-learning–based-deception detection techniques and other nonmachine-learning-based–deception detection techniques [58]. The first stream of literature is built on tenets of design science—specifically using machine-learning techniques and verbal and nonverbal features—to detect malicious behavior [58]. The machine-learning techniques used in this stream of literature can be further classified into two subcategories: supervised learning and unsupervised learning. The second stream of literature is based on methods outside of the machine-learning context, for example, oculometric analysis, where it is assumed that a person's emotional state is reflected physiologically.

Supervised Machine-Learning Techniques

Lau et al. [37] apply an SVM-based classifier with text-mining methods and semantic language models for the detection of untruthful reviews hosted on amazon.com. Ho et al. [28] examine supervised-learning methods, such as SVM, decision tree, and logistic regression, to detect deception in spontaneous online communication. They find that deceivers typically use fewer words than truth-tellers in spontaneous synchronous communication environments. Ludwig et al. [47] explore features of e-mails from business partners participating in deceitful practices to predict deception using an ordinal multilevel regression model. Benjamin et al. [8] investigate cybercriminal communities to identify potential long-term and key participants. Zhang et al. [83] use key reviewer-based features with supervised-learning techniques to detect fake reviews. Siering et al. [66] use a machine-learning approach to examine deception in crowdfunding projects via various linguistic features, such as complexity, diversity, and expressivity, in the textual information.

The problem of detecting deceptive reviews and reviewers using supervised-learning methods has been studied in the computer science (CS) domain as well. Dave et al. [15] use a number of supervised-learning methods for review classification and compare their effectiveness on whole reviews versus individual sentences. Jindal and Liu [32] experiment with a supervised-learning approach that utilizes

features derived from behavior analysis of spammers to identify opinion spam in a data set containing product reviews from amazon.com. Li et al. [39] use a supervised-learning method with a *co-training* method to detect spammers based on review features, and spam reviews based on spammer features. Lin et al. [44] conduct experiments using supervised-learning techniques in conjunction with threshold-based solutions to detect fake reviews. They propose six time-sensitive features to detect fake reviews as early as possible. Li et al. [40] employ a feature-based sparse additive generative model as well as SVM in their classification to explore the general rule for deceptive opinion spam detection.

Our study is most similar to this stream of literature because we focus on several machine-learning techniques. However, instead of using raw or derived features, we focus on feature engineering and incorporate several new features based on their distributional aspects.

Unsupervised Machine-Learning Methods

Unsupervised-learning methods do not use labeled data to detect or classify spam and genuine reviewers. Liang et al. [42] employ unsupervised text-mining techniques to identify common characteristics of a large number of malicious insiders. Lim et al. [43] develop a comprehensive scoring method to measure the degree of overall spam behavior of each reviewer. Wu et al. [80] assess the trustworthiness of reviews based on the distortion of product ranking in the absence of gold-standard data. The underlying assumption is that spam reviews distort the product ranking more than genuine reviews.

Other (Nonmachine-Learning) Techniques

The problem of deception detection on digital platforms has been studied outside of the machine-learning domain as well. Proudfoot et al. [61] show that deception detection systems can employ oculometric behaviors, such as pupil dilation and eye-gaze fixation, to improve performance. Lappas et al. [36] study the vulnerability of individual businesses to online fake review attacks. Akoglu et al. [4] and Rayana and Akoglu [63] use a probabilistic graphical model to detect opinion spammers in restaurant reviews from yelp.com. Ye and Akoglu [81] use abnormal network footprints as features to detect coordinated groups of spammers. Wang et al. [74] adopt a graph-based approach and investigate the relationship between nodes associated with reviewers, reviews, and stores to detect online store reviewer spammers.

Our study differs from this stream of research by focusing on deception detection methods with supervised-learning techniques in a hierarchical manner using a new feature engineering approach.

Feature Engineering in Deception Detection

Machine-learning and other (nonmachine-learning) methods of detecting deception require input features to develop models. The selection of appropriate features plays a critical role in the performance of these models. Zhang et al. [83] classify input features into two categories: verbal (features extracted from review text) and nonverbal features of review-posting behavior (social interactions with other reviewers). They highlight the effectiveness of using nonverbal features, such as the number of friends and the number of local photos taken, to detect online fake reviews. Feng et al. [19] use syntactic stylometry features for deception detection. They find statistical evidence of deep syntactic patterns in the data set and demonstrate that features driven from context-free grammar (CFG) parse trees improve deception detection performance. Banerjee and Chua [7] adopt linguistic cues such as readability, genre, and writing style of negative reviews to build key features and train logistic regression to identify manipulative or authentic reviews. Duan and Zirn [17] use various textual and nontextual features to detect suspicious reviews, reviewers, and objects of the reviews.

To the best of our knowledge, the use of the distributional aspect to exploit features in deception detection is a promising area of research and has not been examined extensively in the literature.

Data Description

Our data set consists of restaurant reviews (yelp.com) used by Rayana and Akoglu [63]. Yelp, founded in 2004 as a third-party, review-hosting social media platform, is a very prominent restaurant review platform, which is why we have chosen it for this study. Yelp allows consumers to provide reviews and ratings of local businesses in order to share their experiences. By the end of the first quarter of 2016, it had hosted more than 102 million reviews [82]. In that first quarter of 2016, it had a monthly average of 21 million unique visitors who visited the app and 69 million unique visitors who visited the site via mobile web [82].

The data set we have used in this study includes the review rating score, the date it was posted, the user who posted it, and the restaurant to which it refers. The data set has 5,044

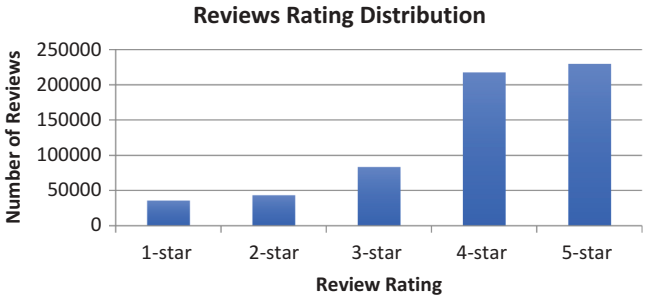

Figure 2. Review Rating Distribution

restaurants from the states of New Jersey, Pennsylvania, Virginia, and Connecticut. It has 260,277 users who wrote reviews between July 2010 and November 2014. Every review has a star rating (a value of 1 to 5, where 1 means least satisfied and 5 means most satisfied). Figure 2 shows a distribution of the review ratings.

We observe that 35,600 (~5.8 percent) reviews have 1-star ratings, 42,985 (~7.1 percent) reviews have 2-star ratings, 83,139 (~13.7 percent) reviews have 3-star ratings, 217,465 (~35.7 percent) reviews have 4-star ratings, and 229,409 (~37.7 percent) reviews have 5-star ratings. Various research projects in the literature have reported results consistent with the overwhelmingly favorable response represented by the J-shaped distribution observed in our sample data [31]. This suggests that our sample is representative of the typical review population.

Reviewer Features Generation

To consider various aspects of reviewing behavior in modeling opinion spammers' behavior, first we generate various univariate and multivariate user-reviewing features. In this section, we describe these univariate and bivariate features that we use in our study.

Univariate Features

We adopt the following features in our analysis: review gap, review count, rating entropy, rating deviation, time of review, and user tenure. These univariate features are not new and have been used in previous studies (see Table 1) for detecting opinion spam. The main contribution of our study is the distribution-based transformation of univariate and bivariate features, which we discuss in detail later.

Review Gap

The review gap between the messages is a convenient statistic to help identify potential spammers who are likely to try to mimic the uniform restaurant-attendance

Table 1. Features Used in the Prior Studies

Features	Used in the Prior Studies
Review Gap	Fei et al. [18], Mukherjee et al. [55, 57]
Review Count	Zhang et al. [84], Wang et al. [76], Mukherjee et al. [55, 57], Luca and Zervas [46]
Rating Entropy	Ye and Akoglu et al. [81], Mukherjee et al. [55, 57]
Rating Deviation	Jindal and Liu [32], Lim et al. [43], Fei et al. [18], Mukherjee et al. [57]
Time of Review	Lim et al. [43], Mukherjee et al. [56], Mukherjee et al. [55, 57]
User Tenure	Goes et al. [25], Ma et al. [49], Wasko et al. [78], Khansa et al. [33]

schedule as well as the review-writing frequency and gap of the average person. Some users post messages in bursts, whereas other users' posting behavior exhibits a more uniform interval between posts. Mukherjee et al. [56] demonstrate that opinion spammers are usually not longtime members of a site. Genuine reviewers, however, use their accounts from time to time to post reviews. Therefore, if reviews are posted over a relatively long timeframe, it suggests normal activity. However, when all reviews are posted within a short burst, it indicates suspicious behavior [55, 57]. We model this gap (in days) between successive reviews as follows:

$$G_i = \frac{1}{N_i - 1} \sum_{j=2}^{N_i} \left(T_{i,j} - T_{i,j-1} \right),$$

where G_i corresponds to the review gap for the ith user, N_i is the number of reviews written by the ith user, and $T_{i,j}$ corresponds to the timestamp of the jth review for user i. Figure 3(a) displays the empirical distribution for review gap (in days) over all users.

Review Count

The number of reviews written by a particular person is a valuable factor in helping distinguish between fake and genuine users. Paid users may generate more reviews than unpaid users. In other cases, a spammer could post very few reviews from one account and create a new account to avoid being detected or blacklisted. The previous literature on review manipulation shows that opinion spammers have behavioral distributions different from nonspammers, and, in particular, posting many reviews indicates abnormal behavior [55, 57]. Therefore, we consider the distribution of the total number of reviews written by a user. Figure 3(b) shows the empirical distribution corresponding to the review count or number of reviews written by each user. We can observe that the distribution is highly skewed in nature.

Rating Entropy

The genuine reviewer tends to base similar reviews on merit, so they may be balanced, that is, equally critical or noncritical in nature. By contrast, we expect spammers to post extreme reviews since their goal is either to artificially improve a particular restaurant's rating or to bring a bad reputation to its competitors. Luca and Zervas [46] empirically find that suspicious reviews tend to be more extreme than normal ones. Mukherjee et al. [55, 57] argue that spammers are more likely to give extreme ratings. To measure this balance or randomness, we calculate the entropy of a particular user's rating scores. We model the entropy E_i of a given user's ratings as follows:

$$E_i = \sum_{j=1}^{K} p_{i,j} \log\left(p_{i,j} \right),$$

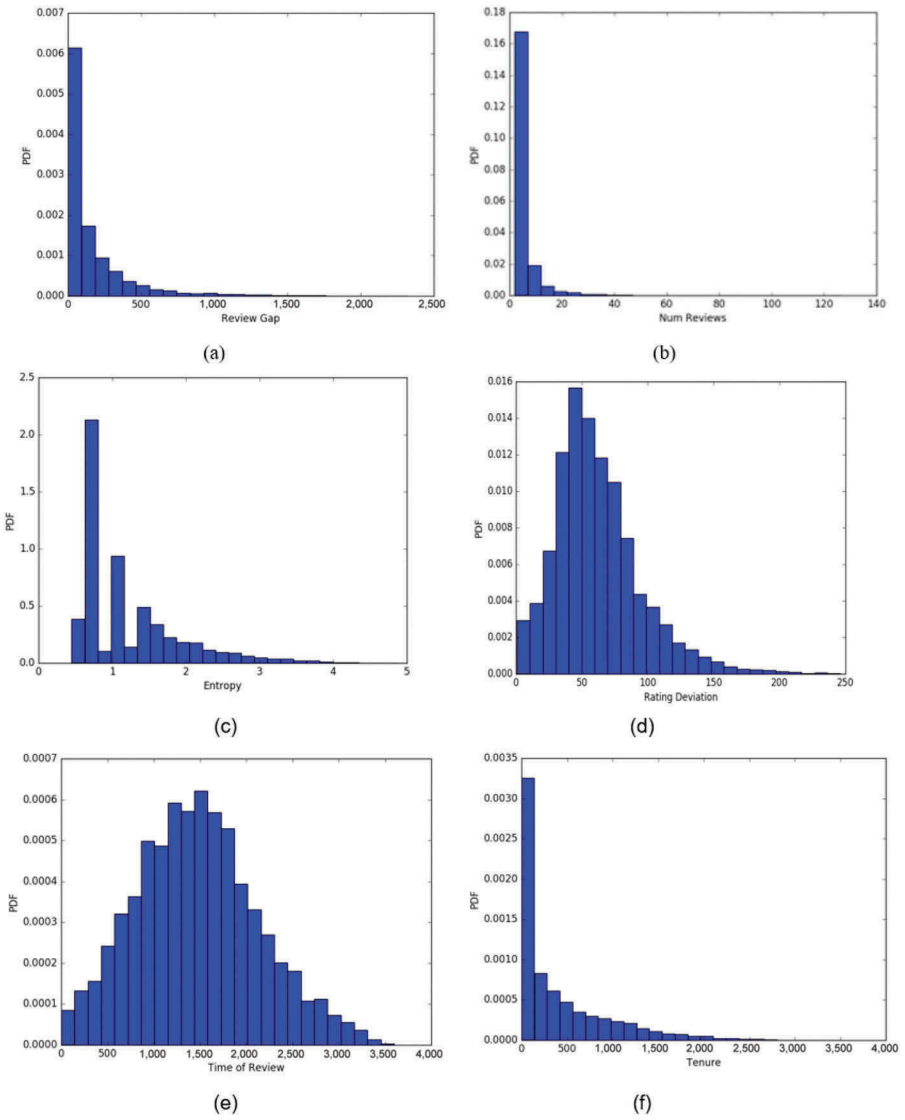

Figure 3. Empirical Distributions of Univariate Features

where $p_{i,j}$ is the probability of user i assigning a review score equal to j, and K is the number of discrete rating scores that can be given by a user. The distribution for rating entropy, which is highly skewed in nature, appears in Figure 3(c).

Rating Deviation

Imagine a spammer who assigns a low rating to every restaurant without regard for other ratings. This type of spammer certainly exists, so we need to take into account

a user's deviation from the average restaurant rating. Therefore, if genuine users who review restaurants fairly outnumber spammers, we can detect those instances where a user's rating deviates greatly from those of other users. However, a bad-dining experience could cause a genuine user to deviate from other users, too. But, if we take the average of deviations across all restaurants reviewed by a user, then it is unlikely that a genuine user's opinion will vastly differ from the majority of users in every case. Lim et al. [43] observe that spammers are likely to deviate from the general rating consensus. Thus, we compute the mean absolute deviation of each user from the average rating of all restaurants reviewed by that user, and we formalize this intuition as follows:

$$D_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\left|R_{i,j} - \mu_{H(j)}\right|,$$

where D_i corresponds to the rating deviation of the ith user, N_i is the number of reviews written by the ith user, $R_{i,j}$ is the rating score given by the ith user in his/her jth review, which corresponds to restaurant $H(j)$, and $\mu_{H(j)}$ is the mean rating for this restaurant. Figure 3(d) shows the empirical distribution for this feature.

Time of Review

This univariate feature models the position of a user in the timeline of restaurant reviews. One strategy spammers may use is to post extremely early after a restaurant's opening in order to maximize the impact of their review. For this reason, it should be a red flag if we notice a user who always posts restaurant reviews before any other user. Lim et al. [43] and Mukherjee et al. [55, 57] argue that early reviews can greatly impact a consumers' sentiment on a product and, in turn, impact sales, so it makes sense for spammers to try to write reviews early.

In the context of restaurant reviews, Luca and Zervas [46] find that review manipulation is driven by economic incentives: Restaurants are more likely to engage in positive review manipulation when they have fewer reviews. The underlying logic is similar: As a restaurant receives more and more reviews, the marginal benefit of each additional positive review decreases. In other words, having more reviews reduces incentives for positive review manipulation. More broadly, Branco and Villas-Boas [10] show that market participants whose survival depends on their early performance are more likely to break rules as they enter the market. Here, we model this phenomenon using the average time (in days) difference between when a review is posted and when the very first review of that same restaurant is posted. Specifically, for the ith user, this feature, that is, average time of review Z_i is represented as follows:

$$Z_i = \frac{1}{N_i}\sum_{j=1}^{N}(T_{i,j} - D_{H(j)}),$$

where N_i is the number of reviews written by the ith user, $T_{i,j}$ is the timestamp for the jth review written by the ith user, $H(j)$ is the restaurant corresponding to the jth review, and $D_{H(j)}$ is the timestamp for the initial review for this restaurant. Figure 3 (e) represents this feature's distribution over all users.

User Tenure

User tenure is defined as the amount of time a reviewer is active in the online forum [25, 33 49, 78]. The confidence regarding the authenticity of a reviewer increases with the amount of time the reviewer is active in the forum. Fake reviewers tend to have short-lived accounts characterized by relatively high volume of reviews and handles, usernames, or aliases designed to avoid detection via pattern recognition algorithms [18]. We therefore use the time period (number of days) in which the user is active as a feature to recognize fake reviewers. Specifically, this feature Y_i is represented as:

$$Y_i = T_{i,N_i} - T_{i,0},$$

where $T_{i,0}$ is the timestamp for the first review written by the ith user, and T_{i,N_i} is the timestamp for the last review written by the ith user. Figure 3(f) shows the empirical distribution corresponding to this feature. As seen here, the distribution is highly skewed since fewer users tend to be active over longer time periods.

Features Summary Statistics

Data preprocessing plays a critical role in the predictive accuracy of a classifier. Following Lim et al. [43], which shows the necessity of preprocessing the data in detecting spammers, we remove inactive reviewers in our data set. More specifically, we exclude (inactive) reviewers who have written fewer than two reviews. Also, we filter restaurants that have less than three reviews in our data set spanning from July 2010 to November 2014. In Table 2, we provide the summary statistics of the preprocessed features used in this study.

Table 2. Summary Statistics of Features

Variables	Min	Max	Mean	SD
Review Gap	0	1,594	140.07	211.08
Review Count	2	165	4.64	6.43
Rating Entropy	0.45	5.08	1.17	0.69
Rating Deviation	0	3.3	0.82	0.44
Time of Review	0	1,611	745.7	372.14
User Tenure	0	1,613	328.37	393.66

Multivariate Features

So far, we have only modeled reviewer behavior using a single variable. We now consider the joint dependency across a user's rating scores. Specifically, we consider the multivariate distribution defined over the rating scores of users. We model each user's set of rating scores as a categorical distribution, where each discrete rating score represents a single category, and then learn a Dirichlet distribution to generalize the categorical distributions over all users.

The Dirichlet distribution is a conjugate prior for the categorical distribution and is therefore a natural choice for our multivariate model. Let u_i be a categorical distribution for the ith user. In our case, we use five categories since a user can rate a restaurant anywhere between 1 and 5 stars. The elements of u_i correspond to the probability of the user's rating score being ≤ 1; between 1 and 2; between 2 and 3; between 3 and 4; and between 4 and 5. For example, suppose a user has reviewed 10 restaurants and assigned a rating of 1-star to 5 restaurants and a rating of 5-stars to the remaining ones, then $u_i = [0.5, 0, 0, 0, 0.5]$. We define a Dirichlet distribution over the categorical distributions of all users. The Dirichlet distribution is parameterized by a set of concentration parameters $\boldsymbol{\alpha}$, and the joint distribution is given by the following equation:

$$P(\mathrm{U}|\boldsymbol{\alpha}) = \frac{\Gamma \sum_k \alpha_k}{\prod_k \Gamma \alpha_k} \prod_k P_k^{\alpha_k - 1},$$

where U represents a user, P_k is the probability that a user assigns a rating score equal to k for a review, α_k is the kth component of $\boldsymbol{\alpha}$, and $\sum_{k=1}^{K} P_k = 1$, where K is the total number of categories.

We estimate the concentration parameters of the Dirichlet distribution (α) from data $\mathcal{D}$, where each instance is a categorical distribution corresponding to a single user. Let $\mathcal{D} = \{u_i\}_{i=1}^{M}$ be the training set where each u_i is a 5-dimensional categorical distribution with $p_{i,k}$ as the kth dimension of this distribution and is equal to the probability that the ith user assigns a rating score between $k-1$ and k. We estimate the concentration parameters of the Dirichlet distribution by maximizing the log-likelihood over $\mathcal{D}$. The log-likelihood of $\mathcal{D}$ is given by:

$$\begin{aligned}\log \mathrm{P}(\mathcal{D}|\boldsymbol{\alpha}) &= \sum_i \log \mathrm{P}(u_i|\boldsymbol{\alpha}) \\ &= M \log \Gamma\left(\sum_k \alpha_k\right) - M \sum_k \log \Gamma(\alpha_k) + M \sum_k (\alpha_k - 1) \log p_k,\end{aligned}$$

where $p_k = 1/\mathrm{M} \sum_i \log p_{i,k}$.

We compute the concentration parameters $\boldsymbol{\alpha}$ that maximize the likelihood function $\log \mathrm{P}(\mathcal{D}|\boldsymbol{\alpha})$ in the above equation. We randomly initialize parameters $\boldsymbol{\alpha}$ and update the concentration parameters in each iteration until we converge to a fixed point that can be shown to be a global maximum for the log-likelihood function corresponding

to the Dirichlet distribution [53]. The updated equation for the concentration parameters is represented as follows:

$$\Psi\left(\alpha_k^{t+1}\right) = \Psi\left(\sum_{\mathrm{k}} \alpha_k^t\right) + \log p_k,$$

where Ψ is the digamma function, and α_k^t is the value of parameter α_k in iteration t. The digamma function is inverted in each iteration to compute the concentration parameters for the next iteration.

Features Engineering Approach

Developing and incorporating various univariate and multivariate user-reviewing features and distributional aspects to detect opinion spammers is one of the main contributions of this study. We assume that even though individual reviews by a spammer may look genuine, collectively we can capture anomalies in the review patterns by modeling user characteristics and interactions among them as univariate and multivariate distributions.

We are hypothesizing that the proposed new feature engineering approach of using distributional characteristics in conjunction with supervised-learning techniques increases the accuracy of detecting spammers over the traditional approach, which ignores distributional aspects of features. We have used reviewers' features with their distributional characteristics to demonstrate the efficacy of our approach. Specifically, we seek to answer the following questions: (1) Will using the distributions corresponding to features have better performance as compared to simply using the standardized feature values within a classifier? (2) What is the specific contribution of each feature toward characterizing overall spammer behavior?

Note that each of the features reviewed in the section titled Reviewer Features Generation can be used, as is, within any supervised-learning algorithm such as logistic regression, naive Bayes, and k-nearest neighbors, to classify opinion spammers. In fact, this method is used in several existing studies, such as Aggarwal et al. [3]. However, this is typically not an ideal approach because the scale of each feature may be different. This may bias the classifier by giving more importance to those features that have larger values while diminishing the effect of features with smaller values. To address the problem of features being noncommensurate, it is common to preprocess the data to make the feature values have a common range. The simplest preprocessing methods are: (1) Linearly rescaling each feature to a common range of numeric values such as [−1,1], and (2) standardizing the data, that is, each feature value is subtracted from its mean and divided by the standard deviation. However, the problem with linear scaling or standardizing the data is that if the original feature values are highly skewed, then the difference between feature values might become less amplified in the transformed feature space.

To overcome these problems, we propose a novel approach where we transform each feature, univariate or multivariate in nature, according to its underlying

probability distribution. The main idea is to generalize the features better with the help of distribution-based transformations. This is particularly important when we want to leverage patterns in skewed data, which is the case for most of the review features that we consider for our task. The transformation itself is done in two steps.

In step 1, we model each feature by fitting it with a best distribution after trying several distributions and choosing the one with the minimum mean squared error (MSE). The main idea behind this step is to better generalize the feature by minimizing noise in the feature. Without this, perturbations in the signal given by the feature can mislead the final classifier. We are essentially smoothing the feature values by explaining them with a standard distribution family.

In step 2, given a distribution for a feature (computed in step 1), we perform a transformation on the probability density function (PDF) values for that feature. It is easy to work with a simple functional form (linear in our case) in a transformed variable instead of using a complex form in the original space [9]. We choose transformations, such as log-linear, and then use least squares to fit a linear function through the transformed distribution. Thus, for each feature f, we obtain a closed-form representation for the feature, $\phi(f)$, given by the parameters of the linear function. This step helps induce linear patterns in each feature, which may not otherwise be seen in the nontransformed space. Specifically, the transformation often helps bring extreme data points with respect to the feature distribution closer to each other as compared to the original nontransformed PDF of the feature. Since the extreme points in the distribution typically represent the anomalous (or spammer) data in each of our features (e.g., number of ratings written, gap between ratings etc.), the transformation may represent the true relationship reasonably well [9] and induce a more discernible pattern for such users. This pattern in the features helps in improving the accuracy of a classifier.

Our approach is generic and can be used to identify fake reviews from any domain by using distributional aspects of relevant univariate and multivariate features. The generality of our approach is the rationale for replacing feature f with $\phi(f)$. The traditional approach of using X can be regarded as a special case by assuming $\phi(f) = f$. However, in our approach, we are choosing the optimal functional form of $\phi()$, in terms of some performance metrics.

Our feature transformation approach can be regarded as an adaptation of a well-known machine-learning technique called "stacking" that combines different classifiers. Specifically, Wolpert [79] proposes a technique where the idea is to learn separate heterogeneous base-classifiers and then combine them together using a second-level classifier. Ting and Witten [70] apply this general strategy to learn probability distributions corresponding to the naive Bayes classifier, and then use other supervised classification methods (such as logistic regression) to combine the naive Bayes distributions together. Here, we further generalize this approach since our new feature space is now a set of different (transformed) probability distributions. Note that each of the transformed features can be independently used as a classifier using a threshold, that is, an instance $\boldsymbol{X}$ can be labeled as a spammer if $\mathrm{P_i}(\boldsymbol{X}) \leq \beta_\mathrm{i}$. Thus, transforming from the original features to the new features has the same effect as stacking multiple base-classifiers

Table 3. Performance Evaluation (MSE) of Spammers Features Distributions and Transformations

Feature	Log-Lin	Log-Log	Power	Exp	Norm	Lognorm	Dirichlet
Review Gap	0.71	**0.48**	123.91	73.52	X	X	X
Review Count	1.13	**0.17**	82.83	246.68	X	X	X
Rating Entropy	**0.44**	0.89	3.93	2.71	4.28	2.36	X
Rating Deviation	**0.57**	1.76	45.53	23.26	18.29	13.78	X
Time of Review	**0.86**	1.21	75.39	38.12	25.44	19.09	X
User Tenure	**0.13**	0.60	90.08	45.22	32.06	24.53	X
Joint Model (User-Rating)	*	*	*	*	*	*	−3.23

X—The error was extremely high for these cases and is not shown.
*Not applicable.

together. As in stacking, we combine the base-classifiers using a second-level classifier, for which we utilize several supervised machine-learning algorithms.

We find the best-fit distribution for each univariate feature generated in the section titled Reviewer Features Generation. We use MSE as a performance measure for the best-fit distribution and transformation. The MSE is measured using the predicted probability and the observed probability on the unseen data. The average MSE for each univariate feature using fivefold cross validation is shown in Table 3. For example, the log-log transformation is found to be the best fit for the review count feature (lowest MSE: 0.17). For multivariate feature, we compute the average log-likelihood to estimate the fit of the multivariate distribution (Joint model—user-rating) and the results are shown in the last column of Table 3.

Model Development and Evaluation

Building a supervised-learning model and evaluating its performance in terms of opinion spammer detection is a challenging problem. One of the issues of building a machine-learning algorithm to detect an opinion spammer is the lack of labeled or annotated data. As a prerequisite to build a pertinent supervised machine-learning model, we need to observe and label each review as fake or genuine. Different approaches have been used in the literature to label reviews, from using Amazon Mechanical Turk (AMT) to generate fake and nonfake reviews [59, 80] to manually annotating individual reviews [43]. However, fake reviews either manually annotated or generated through AMT tend to be biased in nature and to be not effective [57].

Given these concerns, we follow the approach used by Goswami et al. [26], Luca and Zervas [46], Mukherjee [57], Rahman et al. [62], and Zhou and Duan [86], and rely on the Yelp classification mechanism to categorize a review as fake or genuine for training and test data set. Yelp is the only major digital platform that does not delete filtered reviews [46]. Moreover, its commercial algorithm for filtering reviews

has been in place since 2005 [68]. Though Yelp's filtered reviews may not be a perfect indicator of fake reviews, its usage as a proxy to label and detect fake reviews has been well established in the academic literature [36, 46, 83]. Following this established literature, we use a sample of reviews filtered by Yelp's proprietary algorithm for the purpose of labeling.

Note that Yelp's algorithm of filtering reviews is proprietary in nature. We are not trying to mimic or reverse engineer the algorithm deployed by Yelp, rather our goal is to propose a new approach that increases the likelihood of detecting spammers over the traditional approach. To achieve this goal, we rely on the Yelp classification mechanism to categorize a review as fake or genuine for training and test data set. Also, we do not have access to all the information Yelp does. Yelp can use its internal data, such as features deriving from IP addresses, locations, network/session logs, click behaviors, social network interactions of reviewers at its website, and so forth [57, 77]. Next, we present the supervised-learning models we deploy on the features engineered in the section titled Features Engineering Approach.

Supervised-Learning Models

We evaluate our approach using several supervised-learning classification models. Before applying any supervised-learning models, we compute the correlation coefficients between features (after transformation) to avoid any potential issues due to correlation on the performance of classification models. From the correlation matrix (Table 4), we can see that the correlation issue is not serious, since the magnitude of correlation coefficients is relatively small.

We consider the log probabilities as features within a supervised classifier in order to avoid arithmetic underflow in the case of probability values that are too small to be stored in memory. We refer to this as the level-1 classifier. We use several supervised machine-learning methods for the level-1 classifier. Each method can be regarded as inducing a unique mixture of the underlying probability distributions. Thus, depending on the level-1 classifier used, the mixture can be either a simple linear combination of the probability distributions or a more complex nonlinear combination of the distributions. Below, we briefly describe the types of classifiers that we use in our experiments.

1. Logistic regression: This classifier assumes that the posterior distribution $P(y|\boldsymbol{X})$, where y is the label and $\boldsymbol{X}$ is the set of features, takes the shape of a logistic function. Logistic regression induces a linear classifier, that is, using logistic regression as a level-1 classifier is equivalent to computing our final classifier as a weighted linear combination of the base probability distributions. To avoid overfitting, we use logistic regression with L-2 regularization, which ensures that the weights learned by the logistic regression classifier do not become too large.
2. Naive Bayes: The naive Bayes classifier makes the simplifying assumption that the features are independent given the class. This corresponds to

Table 4. Correlation Coefficients between Features

	Review Gap	Rating Entropy	Review Count	Joint Distribution (User-Rating)	Rating Deviation	Reviewer Tenure	Time of Review
Review Gap	1.000	−0.170	−0.159	−0.243	0.134	−0.161	0.093
Rating Entropy	−0.170	1.000	−0.238	−0.118	−0.173	−0.199	−0.130
Review Count	−0.159	−0.238	1.000	0.180	−0.105	−0.234	−0.186
Joint Distribution (User-Rating)	−0.243	−0.118	0.180	1.000	−0.126	−0.259	0.227
Rating Deviation	0.134	−0.173	−0.105	−0.126	1.000	0.088	0.040
Reviewer Tenure	−0.161	−0.199	−0.234	−0.259	0.088	1.000	−0.215
Time of Review	0.093	−0.130	−0.186	0.227	0.040	−0.215	1.000

assuming that each of our base distributions is conditionally independent of the others, given the class (whether the user is a spammer or not). Note that, even though this assumption does not hold true for real data sets, naive Bayes classifiers tend to perform well in classification tasks, since the actual probability value is not as important as predicting the right class [16]. That is, if the right class has a higher probability value computed by the naive Bayes classifier as compared to the wrong class, it is enough for correct classification. In fact, Domingos and Pazzani [16] show the optimality of naive Bayes under specific conditions.

3. K-nearest neighbors: In k-nearest neighbors, we compute the neighborhood of a point using Euclidean distance and determine the class by taking a majority vote in this neighborhood. In our case, this is equivalent to computing the neighborhood of an example over the transformed feature space.
4. SVM: We use two types of SVMs in our experiments. The SVM-linear uses a linear kernel, and induces a linearly weighted combination of our base distributions. However, the weighting is performed using max-margin learning [13]. The SVM-radial uses a radial kernel, which means that in our case, it induces a nonlinear mixture of the base probability distributions.
5. AdaBoost: AdaBoost is one of the most widely used ensemble models that reduces both bias and variance by combining several weak classifiers [22]. Here, we use AdaBoost with decision stumps as our level-1 classifier.
6. Random forest: Random forest is a type of decision-tree learning, where we learn multiple decision trees and combine them together. This is a form of bootstrap aggregation or bagging for decision trees [11].
7. Classification and regression trees (CART): CART is another type of decision-tree learning that uses an alternative measure to entropy, which is a measure used in the classical ID3 algorithm and its variants, in order to determine which feature to split on while building the decision tree. CART sometimes produces smaller depth trees than trees that use entropy for determining the splitting feature [12].

Results and Discussion

We evaluate the performance of each of the above classifiers using the transformed features. We adopt a standard model evaluation validation approach that is widely used in the IS literature. Following Abbasi et al. [1], Sinha and May [67], Twyman et al. [73] and Zhou and Duan [86], we perform k-fold cross validation ($k = 5$) in our data set and report the final average F1 and area under curve (AUC) scores in our results. In k-fold cross validation, we segment the data into k equal-sized partitions. During each run, we choose one of the partitions for testing, while the remaining k—1 partitions are used for training. We repeat this procedure k times in order to use each partition for testing exactly one time. The average F1 score is computed by

Table 5. Performance Comparison of Classifiers Using Transformed Features

Algorithm	Precision	Recall	F1 Score	AUC Score
Logistic Regression	0.827	0.709	0.763	0.817
k-NN (k = 20)	0.709	0.782	0.744	0.789
k-NN (k = 10)	0.657	0.864	0.746	0.781
k-NN (k = 5)	0.802	0.673	0.732	0.760
Naive Bayes	0.623	0.814	0.706	0.753
AdaBoost	0.710	0.833	0.767	0.732
CART	0.608	0.801	0.691	0.724
Random forest	0.546	0.772	0.640	0.717
SVM (Linear)	0.540	0.757	0.630	0.701
SVM (Radial)	0.545	0.796	0.647	0.712

averaging the F1 scores for all runs. We also compute the AUC for several different tunable hyperparameters (e.g., threshold).

Table 5 shows the results of performing classification using the transformed feature space and each of the classifiers reviewed in the section subtitled Supervised-Learning Models. As seen here, the best performing method is logistic regression, which achieves an AUC score of 0.817. Logistic regression is a probabilistic classifier that assigns probabilities associated with each class. It tends to perform extremely well in the conditions such as: (1) low dimensional features space, that is, the numbers of features are relatively small as compared to the number of observations, and (2) not too many categorical features. These two conditions might have been found favorable in our specific data set, thereby making logistic regression the best-performing algorithm. The k-NN algorithm performs nearly as well as logistic regression when using a k value equal to 20. The performances of naive Bayes, AdaBoosting, and random forest are similar to each other, whereas CART performs slightly worse than these three approaches. SVMs with either radial or linear kernels perform the worst among all the methods we evaluate. SVM

Table 6. Performance Comparison of Classifiers Using Standardized Features

Algorithm	Precision	Recall	F1 Score	AUC Score
Logistic regression	0.574	0.737	0.645	0.723
k-NN (k = 20)	0.559	0.709	0.625	0.701
k-NN (k = 10)	0.606	0.659	0.632	0.688
k-NN (k = 5)	0.694	0.570	0.626	0.676
Naive Bayes	0.630	0.584	0.606	0.668
AdaBoost	0.653	0.542	0.592	0.659
CART	0.674	0.558	0.611	0.657
Random forest	0.616	0.569	0.592	0.651
SVM (Linear)	0.629	0.527	0.573	0.636
SVM (Radial)	0.608	0.577	0.592	0.665

performs well, especially in high dimensional feature space where large numbers of features explain the overall behavior of interest better, which is not the case with respect to the specific data set and features being used in this study.

As mentioned earlier in the section titled Features Engineering Approach, an alternative approach to using the user features is to use the standardized value for a feature directly in a classifier. We next present the results of classification when using the standardized features. Corresponding to each of our univariate distributions, we have one standardized feature and, corresponding to our multivariate distribution, we simply use the raw frequency of users' ratings and average restaurant ratings. The results obtained using standardized features are shown in Table 6.

Our approach of using the probability distribution performs much better than using features directly. In each classifier, both the F1 score and AUC score (Table 5) are considerably better than the approach that does not consider the distributional aspect of the features (Table 6). The proposed approach results in improvement of the AUC score by about 13 percent and the F1 score by about 18 percent using the best-performed classifier. These results are quite encouraging as they achieve relatively high AUC and F1 scores using distribution-based features.

Finally, we compare each individual distribution's contribution to the classification of spammers. This experiment allows us to determine the relative importance of each feature distribution in our overall model. We use our best-performing algorithm, in the case of stacked classifiers: logistic regression. We consider each of our distributions independently as a single feature in the logistic regression model to evaluate performance based on a single distribution. Table 7 shows the F1 and AUC scores corresponding to each distribution. This experiment allows us to determine the relative importance of each distribution in our overall model.

As expected, using AUC and F1 scores as the performance measures, we see that stacking the distributions to form a collective classifier performs significantly better than simply relying on a single distribution's results. As shown in Table 7, the average gap (AUC score: 0.6797 and F1 score: 0.7248) plays an important role in detecting spammers, followed by the distributions that model ratings entropy and review count. The joint distribution has the next best performance (in terms of the

Table 7. Contribution of Individual Distributional Characteristics in Spammer Detection

Input	F1 Score	AUC Score
Review Gap Distribution	0.7248	0.6797
Rating Entropy Distribution	0.6692	0.6534
Review Count Distribution	0.6515	0.6473
Joint Distribution (User-Rating)	0.6013	0.6434
Rating Deviation Distribution	0.6247	0.6204
Time of Review Distribution	0.5932	0.6121
Reviewer Tenure Distribution	0.5430	0.5954

Table 8. Performance Comparison of Classifiers Using Validation Dataset

	Transformed Features		Raw Features	
Algorithm	F1 Score	AUC Score	F1 Score	AUC Score
Logistic Regression	0.7901	0.7686	0.7076	0.6837
SVM (Linear)	0.7861	0.7598	0.7196	0.6758
k-NN (k = 10)	0.7381	0.7533	0.6827	0.6519
k-NN (k = 5)	0.7261	0.7524	0.6974	0.6377
Naive Bayes	0.7255	0.7333	0.7045	0.6157
Random forest	0.7652	0.7120	0.7132	0.6055
CART	0.7021	0.6941	0.6435	0.5702
AdaBoost	0.6320	0.6712	0.5678	0.5893
SVM (Radial)	0.7759	0.6668	0.6543	0.6193
k-NN (k = 20)	0.6877	0.6455	0.7451	0.6092

AUC score), indicating to an extent the role of the dependency between a user's rating and a restaurant's overall rating in characterizing spammers' reviewing habits. The time of review and reviewer tenure, though important, have the least contribution (in terms of AUC and F1 scores) toward classification performance among all the distributions we consider.

Methodology Validation

We validate our methodology using another data set from a different domain. We collect an additional data set (1,235 reviews) from Yelp with the full history of review information for each post, including review date and review rating, for each hospital in the cities of Memphis and Seattle. We perform the same steps as described earlier on the restaurant reviews. Table 8 shows the classification performance of the models using the transformed features on the hospital review data set.

The results using this new data set are consistent with those based on our original restaurant review data set: All of the classifiers using distribution-based feature-engineering approach perform better than the standardized data in both restaurant review and hospital review data sets (except the F1 score in the last row of Table 8). More specifically, using features with the proposed approach improves the AUC score by about 12.4 percent and F1 score by about 11.6 percent for the best-performing classifier.

Performance Comparison with Prior Studies

There have been studies in the past that have used review data from Yelp, TripAdvisor, Amazon, and so on to detect opinion spam. However, annotating online reviews is a challenging problem since studies have shown that even

human annotators are just slightly better than random, when it comes to classifying reviews as spam or not-spam effectively [60]. Thus, it is possible that results based on such data sets are biased. We compare the results obtained using our methodology with those produced by a few other notable studies conducted in the past. Jindal and Liu [32] evaluate their supervised-learning based approach on an Amazon data set, which they annotate by considering duplicate reviews as fake reviews. Their results show an AUC score of 78 percent for spam (not spammer) detection. Feng et al. [19] use multiple data sets from TripAdvisor and Yelp to evaluate their system, which is based on using probabilistic context-free grammars derived from the review text. Each of these data sets is quite small (around 800 reviews). The TripAdvisor data set is annotated using heuristics, while for the Yelp data set, they rely on the labels provided by Yelp. Again, the focus here is on review spam detection, and the best score they obtain is around 64.3 percent accuracy for the Yelp data set. Rayana and Akoglu [63] evaluate their semisupervised algorithm for identifying fake reviewers based on a combination of metadata and text features and obtain an AUC score of 68.28 percent. Our approach of detecting fake reviewers achieves an AUC score of 0.817 or 81.7 percent and F1 score of 0.763 or 76.3 percent. Clearly, our approach outperforms the other approaches, thereby illustrating the significance of the proposed method.

Conclusions

The issue of opinion spamming in online reviews is not going away, and detecting the perpetrators is not easy. Various approaches have been proposed to discern between insidious spammers and legitimate reviewers, but these approaches do not fully realize the potential of incorporating underlying distributional aspects of review behavior. They typically tend to use features directly in a classifier, which means that the classification is less robust to noise in the features. Also, these approaches do not consider multivariate distributions over users' ratings. In this study, we propose a novel approach on spammer detection where we fit univariate and multivariate distributions to key features derived from the metadata of users that help distinguish spammers from genuine users. We then stack the probability output from the individual distributions into a meta-classifier and empirically show that this approach generalizes more effectively on unseen data than typical approaches that utilize standardized features without considering their underlying distribution. Therefore, our approach of combining univariate distributions with a multivariate distribution across users' ratings yields a more powerful model to detect spammers.

Businesses must keep an eye on digital platforms to ensure reviewers are posting authentic reviews, but this may be very difficult to do manually. One effective spam-detection method that platforms can use is to identify spammers and post their information on the dashboard of the businesses. Such features would make the review-hosting site more attractive to business owners, review writers, and readers. Our proposed approach can help digital platforms (restaurant review platforms, hotel review platforms, movie review platforms, etc.) detect fake reviewers more

precisely. Hence, our approach increases the information credibility of online product and service reviews.

Our results have important managerial implications for online platforms that are prone to increased abuse and manipulation from strategic parties [35, 50]. First, our analysis demonstrates that different features have varying levels of influence on the ability to detect spammers. With access to an ordered list of features, platform owners and businesses (to a certain extent) can prioritize and learn about the most influential features of spammers and then incorporate appropriate interventions as a defensive mechanism against those spammers. Since our methodology provides information about the probability of being a spammer based on the key features, digital platforms can develop a spam score for each reviewer and share it with business owners and consumers. Second, our methodology is especially helpful to firms that want to implement a detection process on their platforms but cannot afford to pour a large amount of resources or time into developing a complex analytical solution. Thus, our easy-to-implement methodology increases the information credibility of online product and service reviews by providing firms a unified framework to detect opinion spammers, which can be subsequently tagged or filtered. Third, from a deployment perspective, this methodology allows for a great degree of scalability, generalization, and customization to the size of any data set and feature complexity, and it offers great insights into both individual and joint features of detection. Potentially, firms can design a fast real-time system using distributional information of multiple features to detect online fake reviews. Our methodology to detect spammers can also be applied in other areas, such as financial fraud detection.

From a business perspective, managers or marketers can take advantage of our design artifacts, using them to detect and remove fake reviews of their products and services. Doing so would enable more effective marketing strategies based on the sheer volume of genuine, user-contributed consumer reviews. At a consumer level, this research would enable consumers to have a better online shopping experience and browse genuine product reviews.

There are several possible extensions to our research. For example, one could crawl the Yelp website to obtain more relevant features (e.g., user's friends, user location, and restaurant location). We expect our approach to work seamlessly with new features and more sophisticated meta-classification methods. Furthermore, we have not utilized any text-based features. We assume that, unlike in e-mail spam, the text would be more deceptive and contextually relevant. Therefore, simplistic text analytics (e.g., bag of words model) would likely have a negative impact on the classifier. However, future research could use more advanced linguistics-based measures to develop adaptive versions of our method. Lastly, we realize that we have not explicitly modeled the underlying economic incentives of conducting review manipulation. As a future research direction, it would be interesting to examine a structural model incorporating the economic incentives to conduct positive or negative review manipulation.

NOTES

1. These online platforms include e-commerce websites such as Amazon; social media sites such as Facebook, Twitter, and Foursquare; and recommendation and review websites such as Yelp, TripAdvisor, and Expedia.

2. See http://usa.chinadaily.com.cn/life/2012–12/13/content_16013662.htm (accessed on August 9, 2017).

3. See http://www.technologyreview.com/view/426174/undercover-researchers-expose-chinese-internet-water-army/ (accessed on August 9, 2017).

4. See http://www.clearmymail.com/guides/viagra_spam_emails.aspx (accessed on August 10, 2017).

5. For readability, we use the term "spammer" to mean opinion spammer.

ORCID


Liangfei Qiu http://orcid.org/0000-0002-8771-9389

REFERENCES

1. Abbasi, A.; Zahedi., F.M.; Zeng, D.; Chen, Y.; Chen, H.; and Nunamaker, J.F. Jr. Enhancing predictive analytics for anti-phishing by exploiting website genre information. *Journal of Management Information Systems*, *31*, 4 (2015), 109–157. doi:10.1080/07421222.2014.1001260

2. Adomavicius, G.; Bockstedt, J.C.; Curley, S.P.; and Zhang, J. Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*, *24*, 4 (2013), 956–975. doi:10.1287/isre.2013.0497

3. Aggarwal, R., and Singh, H. Differential influence of blogs across different stages of decision making: The case of venture capitalists. *MIS Quarterly*, *37*, 4 (2013), 1093–1112. doi:10.25300/MISQ

4. Akoglu, L.; Chandy, R.; and Faloutsos, C. Opinion fraud detection in online reviews by network effects. *Proceedings of the International AAAI Conference on Weblogs and Social Media*, *7*, (2013), 2–11.

5. Anderson, M., and Magruder, J. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *Economic Journal*, *122*, 563 (2012), 957–989. doi:10.1111/ecoj.2012.122.issue-563

6. Associated Press. Fake online reviews: Here are some tips for detecting them. *NBC News*. 2015. http://www.nbcnews.com/business/consumer/fake-online-reviews-here-are-some-tips-detecting-them-n447681 (accessed on August 10, 2017).

7. Banerjee, S., and Chua, A.Y.K. A study of manipulative and authentic negative reviews. *Proceedings of the International Conference on Ubiquitous Information Management and Communication*, *8*, (2014), 1–6.

8. Benjamin, V.; Zhang, B.; Nunamaker, J.F. Jr; and Chen, H. Examining hacker participation length in cybercriminal Internet-relay-chat communities. *Journal of Management Information Systems*, *33*, 2 (2016), 482–510. doi:10.1080/07421222.2016.1205918

9. Box, G.E., and Cox, D.R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological), 26, 2* (1964), 211–252.

10. Branco, F., and Villas-Boas, J.M. Competitive vices. *Journal of Marketing Research*, *52*, 6 (2015), 801–816. doi:10.1509/jmr.13.0051

11. Breiman, L. Random forests. *Machine Learning*, *45*, 1 (2001), 5–32. doi:10.1023/A:1010933404324

12. Breiman, L.; Friedman, J.; Stone, C.J.; and Olshen, R.A. *Classification and Regression Trees*. Boca Raton: CRC Press, 1984.

13. Cortes, C., and Vapnik, V. Support-vector networks. *Machine Learning*, *20*, 3 (1995), 273–297. doi:10.1007/BF00994018

14. Dalvi, N.N.; Kumar, R.; and Pang, B. Para "normal" activity: On the distribution of average ratings. *Proceedings of International Conference on Weblogs and Social Media*, *7*, (2013), 110–119.

15. Dave, K.; Lawrence, S.; and Pennock, D.M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the International Conference on World Wide Web*, *12* (2003), 519–528.

16. Domingos, P., and Pazzani, M. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, *29*, 2 (1997), 103–130. doi:10.1023/A:1007413511361

17. Duan, H., and Zirn, C. Can we identify manipulative behavior and the corresponding suspects on review websites using supervised learning? *Lecture Notes in Computer Science*, *7617* (2012), 215–230.

18. Fei, G.; Mukherjee, A.; Liu, B.; Hsu, M.; Castellanos, M.; and Ghosh, R. Exploiting burstiness in reviews for review spammer detection. *Proceedings of the International AAAI Conference on Weblogs and Social Media*, *7* (2013), 175–184.

19. Feng, S.; Banerjee, R.; and Choi, Y. Syntactic stylometry for deception detection. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, *50* (2012), 171–175.

20. Forman, C.; Ghose, A.; and Wiesenfeld, B. Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*, 3 (2008), 291–313. doi:10.1287/isre.1080.0193

21. Freeman, L.L. How to spot fake online reviews. Time, July 22, 2016. http://time.com/money/4362586/fake-online-reviews/ (accessed on August 10, 2017)

22. Freund, Y., and Schapire, R.E. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, *14*, 5 (1999), 771–780.

23. Fuscaldo, D. How to spot fake reviews online. *Fox News*. June 27, 2014. http://www.foxbusiness.com/features/2014/06/27/how-to-spot-fake-online-reviews.html (accessed onAugust 10, 2017).

24. Gallivan, R. Amid fake reviews, consumers are skeptical of social media marketing. *Wall Street Journal*, June 3, 2014. http://blogs.wsj.com/digits/2014/06/03/amid-fake-reviews-consumers-skeptical-of-social-media-marketing/ (accessed on August 10, 2017).

25. Goes, P.B.; Lin, M.; and Au Yeung, C.M. Popularity effect" in user-generated content: Evidence from online product reviews. *Information Systems Research*, *25*, 2 (2014), 222–238. doi:10.1287/isre.2013.0512

26. Goswami, K.; Park, Y.; and Song, C. Impact of reviewer social interaction on online consumer review fraud detection. *Journal of Big Data*, *4*, 1 (2017), 1–19. doi:10.1186/s40537-017-0075-6

27. Gu, B., and Ye, Q. First step in social media: measuring the influence of online management responses on customer satisfaction. *Production and Operations Management*, *23*, 4 (2014), 570–582. doi:10.1111/poms.2014.23.issue-4

28. Ho, S.M.; Hancock, J.T.; Booth, C.; and Liu, X. Computer-mediated deception: Strategies revealed by language-action cues in spontaneous communication. *Journal of Management Information Systems*, *33*, 2 (2016), 393–420. doi:10.1080/07421222.2016.1205924

29. Hu, N.; Bose, I.; Gao, Y.; and Liu, L. Manipulation in digital word-of-mouth: A reality check for book reviews. *Decision Support Systems*, *50*, 3 (2011), 627–635. doi:10.1016/j.dss.2010.08.013

30. Hu, N.; Liu, L.; and Sambamurthy, V. Fraud detection in online consumer reviews. *Decision Support Systems*, *50*, 2 (2011), 614–626. doi:10.1016/j.dss.2010.08.012

31. Hu, N.; Zhang, J.; and Pavlou, P.A. Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, *52*, 10 (2009), 144–147. doi:10.1145/1562764

32. Jindal, N., and Liu, B. Opinion spam and analysis. In *Proceedings of the International Conference on Web Search and Data Mining. New York,*. NY: ACM, 2008, pp. 219–230.

33. Khansa, L.; Ma, X.; Liginlal, D.; and Kim, S.S. Understanding members' active participation in online question-and-answer communities: A theory and empirical analysis. *Journal of Management Information Systems*, *32*, 2 (2015), 162–203. doi:10.1080/07421222.2015.1063293

34. Kumar, N.; Qiu, L.; and Kumar, S. Exit, voice, and response on digital platforms: An empirical investigation of online management response strategies. *Information Systems Research* (2018), Forthcoming.

35. Lahiri, A.; Dewan, R.M.; and Freimer, M. The disruptive effect of open platforms on markets for wireless services. *Journal of Management Information Systems*, *27*, 3 (2010), 81–110. doi:10.2753/MIS0742-1222270304

36. Lappas, T.; Sabnis, G.; and Valkanas, G. The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research*, *27*, 4 (2016), 940–961. doi:10.1287/isre.2016.0674

37. Lau, R.Y.K.; Liao, S.Y.; Kwok, R.C.W.; Xu, K.; Xia, Y.; and Li, Y. Text mining and probabilistic language modeling for online review spam detection. *Transactions on Management Information Systems, ACM*, *2*, 4 (2011), 1–30. doi:10.1145/2070710.2070716

38. Lee, S.; Qiu, L.; and Whinston, A.B. Sentiment manipulation in online platforms and opinion forums: An analysis of movie tweets. Working paper, University of Texas Austin, 2016.

39. Li, F.; Huang, M.; Yang, Y.; and Zhu, X. Learning to identify review spam. *Proceedings of the International Joint Conference on Artificial Intelligence*, *22*, 3 (2011), 2488–2493.

40. Li, J.; Ott, M.; Cardie, C.; and Hovy, E.H. Towards a general rule for identifying deceptive opinion spam. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, *52*, (2014), 1566–1576.

41. Li, X. Could deal promotion improve merchants' online reputations? The moderating role of prior reviews. *Journal of Management Information Systems*, *33*, 1 (2016), 171–201. doi:10.1080/07421222.2016.1172450

42. Liang, N.; Biros, D.P.; and Luse, A. An empirical validation of malicious insider characteristics. *Journal of Management Information Systems*, *33*, 2 (2016), 361–392. doi:10.1080/07421222.2016.1205925

43. Lim, E.P.; Nguyen, V.A.; Jindal, N.; Liu, B.; and Lauw, H.W. Detecting product review spammers using rating behaviors. *Proceedings of the ACM International Conference on Information and Knowledge Management*, *19*, (2010), 939–948.

44. Lin, Y.; Zhu, T.; Wu, H.; Zhang, J.; Wang, X.; and Zhou, A. Towards online anti-opinion spam: spotting fake reviews from the review sequence. In *Proceedings of International Conference on Advances in Social Networks Analysis and Mining*. IEEE Computer Society, 2014, pp. 261–264.

45. Luca, M. Reviews, reputation, and revenue: The case of Yelp.com. 2011. doi:10.2139/ssrn.1928601

46. Luca, M., and Zervas, G. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, *62*, 12 (2016), 3412–3427. doi:10.1287/mnsc.2015.2304

47. Ludwig, S.; Van Laer, T.; De Ruyter, K.; and Friedman, M. Untangling a web of lies: Exploring automated detection of deception in computer-mediated communication. *Journal of Management Information Systems*, *33*, 2 (2016), 511–541. doi:10.1080/07421222.2016.1205927

48. Luo, X., and Zhang, J. How do consumer buzz and traffic in social media marketing predict the value of the firm? *Journal of Management Information Systems*, *30*, 2 (2013), 213–238.

49. Ma, M., and Agarwal, R. Through a glass darkly: information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research*, *18*, 1 (2007), 42–67. doi:10.1287/isre.1070.0113

50. Mantena, R., and Saha, R.L. Co-opetition between differentiated platforms in two-sided markets. *Journal of Management Information Systems*, *29*, 2 (2012), 109–140. doi:10.2753/MIS0742-1222290205

51. Mayzlin, D. Promotional chat on the Internet. *Marketing Science*, *25*, 2 (2006), 155–163. doi:10.1287/mksc.1050.0137

52. Mayzlin, D.; Dover, Y.; and Chevalier, J. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, *104*, 8 (2014), 2421–2455. doi:10.1257/aer.104.8.2421

53. Minka, T. Estimating a Dirichlet distribution. Technical report, MIT, 2000.

54. Mintel. Seven in 10 Americans seek out opinions before making purchases. 2015. http://www.mintel.com/press-centre/social-and-lifestyle/seven-in-10-americans-seek-out-opinions-before-making-purchases (accessed on August 10, 2017).

55. Mukherjee, A.; Kumar, A.; Liu, B.; Wang, J.; Hsu, M.; Castellanos, M.; and Ghosh, R. Spotting opinion spammers using behavioral footprints. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *19* (2013), 632–640.

56. Mukherjee, A.; Liu, B.; and Glance, N. Spotting fake reviewer groups in consumer reviews. *Proceedings of the International Conference on World Wide Web*, *21*, (2012), 191–200.

57. Mukherjee, A.; Venkataraman, V.; Liu, B.; and Glance, N.S. What Yelp fake review filter might be doing? *Proceedings of the International AAAI Conference on Weblogs and Social Media*, *7* (2013), 409–418.

58. Nunamaker, J.F.; Burgoon, J.K.; and Giboney, J.S. Special issue: Information systems for deception detection. *Journal of Management Information Systems*, *33*, 2 (2016), 327–331. doi:10.1080/07421222.2016.1205928

59. Ott, M.; Cardie, C.; and Hancock, J. Estimating the prevalence of deception in online review communities. *Proceedings of the International Conference on World Wide Web*, *21* (2012), 201–210.

60. Ott, M.; Choi, Y.; Cardie, C.; and Hancock, J.T. Finding deceptive opinion spam by any stretch of the imagination. *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, *49*, 1 (2011), 309–319.

61. Proudfoot, J.G.; Jenkins, J.L.; Burgoon, J.K.; and Nunamaker, J.F. Jr. More than meets the eye: How oculometric behaviors evolve over the course of automated deception detection interactions. *Journal of Management Information Systems*, *33*, 2 (2016), 332–360. doi:10.1080/07421222.2016.1205929

62. Rahman, M.; Carbunar, B.; Ballesteros, J.; and Chau, D.H.P. To catch a fake: Curbing deceptive yelp ratings and venues. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *8*, 3 (2015), 147–161. doi:10.1002/sam.11264

63. Rayana, S., and Akoglu, L. Collective opinion spam detection: bridging review networks and metadata. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *21* (2015), 985–994.

64. Roberts, J.J. Amazon sues people who charge $5 for fake reviews. *Fortune Magazine*. October 19, 2015. http://fortune.com/2015/10/19/amazon-fake-reviews/ (accessed on August 10, 2017).

65. Rudolph, S. The impact of online reviews on customers' buying decisions. *Business 2 Community*. 2015. http://www.business2community.com/infographics/impact-online-reviews-customers-buying-decisions-infographic-01280945#iZwM69pSgVKLlH6A.97 (accessed on August 10, 2017).

66. Siering, M.; Koch, J.A.; and Deokar, A.V. Detecting fraudulent behavior on crowd platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems*, *33*, 2 (2016), 421–455. doi:10.1080/07421222.2016.1205930

67. Sinha, A.P., and May, J.H. Evaluating and tuning predictive data mining models using receiver operating characteristic curves. *Journal of Management Information Systems*, *21*, 3 (2004), 249–280. doi:10.1080/07421222.2004.11045815

68. Stoppelman, J. Why Yelp has a review filter. 2009. http://officialblog.yelp.com/2009/10/why-yelp-has-a-review-filter.html (accessed on August 10, 2017).

69. Stritfeld, D. The best book reviews money can buy. *New York Times*. August 26, 2012. http://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html (accessed on August 10, 2017)

70. Ting, K.M., and Witten, I.H. Stacked generalization: When does it work? *Proceedings of the International Joint Conference on Artificial Intelligence*, *15*, 2 (1997), 866–871.

71. Tsikerdekis, M., and Zeadally, S. Online deception in social media. *Communications of the ACM*, *57*, 9 (2014), 72–80. doi:10.1145/2663191

72. Tuttle, B. Amazon lawsuit shows that fake online reviews are a big problem. *Time* October 19, 2015. http://time.com/money/4078632/amazon-fake-online-reviews/ (accessed on August 10, 2017).

73. Twyman, N.W.; Proudfoot, J.G.; Schuetzler, R.M.; Elkins, A.C.; and Derrick, D.C. Robustness of multiple indicators in automated screening systems for deception detection. *Journal of Management Information Systems*, *32*, 4 (2015), 215–245.

74. Wang, G.; Xie, S.; Liu, B.; and Philip, S.Y. Review graph based online store review spammer detection. *Proceedings of the International Conference on Data Mining*, *11* (2011), 1242–1247.

75. Wang, G.; Xie, S.; Liu, B.; and Yu, P.S. Identify online store review spammers via social review graph. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *3*, 4 (2012), 61–82.

76. Wang, Y.; Chan, S.C.F.; Ngai, G.; and Leong, H.V. Quantifying reviewer credibility in online tourism. In *Proceedings of the International Conference on Database and Expert Systems Applications* New York, NY: Springer-Verlag New York, Inc., 2013, pp. 381–395.

77. Wang, Z. Anonymity, social image, and the competition for volunteers: A case study of the online market for reviews. *BE Journal of Economic Analysis and Policy.*Heidelberg: Springer, *10*, 1 (2010), 1–35.

78. Wasko, M., and Faraj, S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, *29*, 1 (2005), 35–57. doi:10.2307/25148667

79. Wolpert, D.H. Stacked generalization. *Neural Networks*, *5*, 2 (1992), 241–259. doi:10.1016/S0893-6080(05)80023-1

80. Wu, G.; Greene, D.; Smyth, B.; and Cunningham, P. Distortion as a validation criterion in the identification of suspicious reviews. *Proceedings of the Workshop on Social Media Analytics*, *1* (2010), 10–13.

81. Ye, J., and Akoglu, L. Discovering opinion spammer groups by network footprints. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Berlin, Heidelberg: Springer, 2015, pp. 267–282.

82. Yelp. 10 things you should know about Yelp. 2016. http://www.yelp.com/about (accessed on August 10, 2017).

83. Zhang, D.; Zhou, L.; Kehoe, J.L.; and Kilic, I.Y. What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, *33*, 2 (2016), 456–481. doi:10.1080/07421222.2016.1205907

84. Zhang, L.; Ma, B.; and Cartwright, D.K. The impact of online user reviews on cameras sales. *European Journal of Marketing*, *47*, 7 (2013), 1115–1128. doi:10.1108/03090561311324237

85. Zhou, R. Muddy waters. December 13, 2012. http://usa.chinadaily.com.cn/life/2012-12/13/content_16013662.htm (accessed on August 10, 2017).

86. Zhou, W., and Duan, W. Do professional reviews affect online user choices through user reviews? An empirical study. *Journal of Management Information Systems*, *33*, 1 (2016), 202–228. doi:10.1080/07421222.2016.1172460