IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Network-Based Approaches To Spam Detection In Online Social Media: A Comprehensive Review

Bipin Kumar Kushwaha[1], Sandeep Kumar Singh[2]

[1]M. Tech Scholar, Dept. of CSE, Saraswati Higher Education & Technical College of Engineering, (AKTU), Varanasi, India

[2]Assistant Professor, Dept. of CSE, Saraswati Higher Education & Technical College of Engineering, (AKTU), Varanasi, India

Abstract— The exponential growth of Online Social Media (OSM) platforms has provided users with dynamic means of communication, content sharing, and interaction. However, this growth has also attracted malicious entities that exploit these platforms for spamming, spreading misinformation, and executing fraudulent activities. Traditional content-based and user-behavioral spam detection techniques often fall short in detecting sophisticated spam patterns. In response, network-based approaches have gained significant traction by leveraging the structural and relational properties of social networks. This comprehensive review explores the spectrum of network-centric methodologies used in spam detection, including graph-based models, community detection algorithms, propagation patterns, and trust networks. The paper categorizes and analyzes the state-of-the-art techniques, evaluates their strengths and limitations, and highlights emerging trends in network science applied to OSM spam detection. Furthermore, it discusses real-world datasets, performance metrics, and the integration of hybrid frameworks combining machine learning with network-based insights. This review serves as a valuable resource for researchers and practitioners aiming to develop robust and scalable spam detection systems in the dynamic landscape of online social platforms.

Keywords: Online Social Media (OSM), Spam Detection, Network-Based Approaches, Graph Theory, Community Detection, Propagation Modeling, Trust Networks, Hybrid Frameworks, Machine Learning, Social Network Analysis

1. INTRODUCTION

The dissemination of information through online social media portals is a significant factor that producers use in their advertising campaigns and customers use to make product and service selections. People these days heavily rely on written reviews to make decisions, with positive and negative reviews encouraging and discouraging them in their product and service selections. We hypothesize that review communities with low signaling costs, like those that make it simple to post a review, and large benefits, like sites with a lot of traffic, will have more deceptive opinion spam than communities with higher signaling costs, like those that make additional requirements for posting reviews, and lower benefits, like low site traffic. It is now common knowledge that user-generated content contains useful data that can be used for a variety of purposes. We focus on product reviews from customers in this paper. We focus primarily on reviewing opinion spam. User feedback on goods and services is abundant in reviews. Before making a purchase decision, potential customers use them to learn what other people think of a product. Product manufacturers also use them to find problems with their products and marketing intelligence information about their rivals.

There has been a growing interest in mining reviews' opinions from academia and industry over the past few years. However, the majority of the current work has been devoted to using data mining and natural language processing to extract and summarize opinions from reviews. Little is had some significant awareness of the qualities of audits and ways of behaving of analysts. Additionally, no study has been reported on the reliability of reviews' opinions. There is no quality control on the Internet, so anyone can write anything. As a result, there are a lot of low-quality reviews and, even worse, review spam. Web page spam and review spam are similar. Web page spam is prevalent in the context of Web search due to the economic and/or public relations value of a page's rank position in a search engine's results. The use of "illegitimate means" to boost the rank positions of some target pages in search engines is known as web page spam. The issue is similar, but also quite distinct, in the context of reviews. For a variety of reasons, reading opinions online is now very common. For instance, if someone wants to buy a product and sees that most of the reviews are positive, they will probably buy it. If most of the reviews are negative, most people will choose another product. Organizations and individuals can reap significant financial benefits and fame from favorable opinions. This gives review/opinion spam good incentives. Written reviews also aid service providers in raising the level of quality of their goods and services. As a result, these reviews are now a big part of a company's success.

While good reviews can help a business, bad reviews can hurt its credibility and cost the company money. Spammers will take advantage of the fact that anyone can post comments pretending to be reviews, creating a tempting opportunity for them to write fake reviews with the intention of misleading users. The sharing feature of social media and the spread of these false reviews online add to their number. When making or canceling a purchase, consumers are increasingly relying on user-generated online reviews. As a result, it appears that businesses and the general public are increasingly concerned about the possibility of publishing deceptive opinion spam|citations reviews that have been purposefully written to sound authentic and deceive the reader. Although this may come as a surprise, very little is known about the actual rate of

deception in online review communities and even less about the factors that may contribute to it.

On the one hand, the pressure to portray businesses, products, and services in a positive light and the relative ease with which reviews can be written could lead one to believe that most online reviews are fake. On the other hand, it could be argued that review websites cannot provide any value unless there is a low rate of deception. In the context of online reviews, the detection of spam has been the primary focus of research. To identify duplicate opinions, Jindal and Liu, for instance, train models with features derived from the reviewer, product, and review text. Using a standard statistical test, manually compare the psychologically relevant linguistic differences between 40 honest and 42 false hotel reviews. These strategies are useful, but they do not address the prevalence of deception in online reviews. In point of fact, empirical and academic studies of the extent of deceptive opinion spam remain scarce. One reason is that it's hard to find trustworthy gold-standard annotations for reviews, such as labels that indicate whether a review is honest (real) or deceptive (fake). One choice for creating highest quality level marks, for instance, is depend on the decisions of human annotators. However, recent research demonstrates that human readers have difficulty identifying deceptive opinion spam; this is particularly the situation while considering the over believing nature of most human adjudicators, a peculiarity alluded to in the mental double dealing writing as a reality predisposition. Recent large meta-analyses demonstrating the inaccuracy of human judgments of deception, with accuracy rates typically close to chance, support the difficulty of identifying fake reviews. Even though self-reports are difficult and costly to obtain, especially in large-scale settings like the internet, it is not surprising that research on estimating the prevalence of deception has generally relied on self-report methods because humans have difficulty recognizing deceptive messages from cues alone. In addition, self-report methods like diaries and large-scale surveys have a number of methodological issues, such as self-deception and social desirability bias. In addition, revealing one's own deception in online reviews is severely discouraged by the possibility of permanent exclusion from a review portal or damage to a company's reputation. According to signaling theory, each review in our situation is a sign of the product's true, unknowable quality; As a result, the objective of consumer reviews is to reduce the inherent information gap between consumers and manufacturers. In a nutshell, the prevalence of deception ought to be proportional to the costs and benefits of fabricating a review, in accordance with a signaling theory approach.

2. RELATED WORK

This section has taken into account the extensive literature review on Spam Detection in Online Social Media Using a Network-Based Framework for Reviews.

E-mail has historically been the focus of research on spam (Drucker et al., 2002), as well as the Internet (Gy ongyi et al., 2004; Ntoulas and other, 2006). According to Jindal and Liu (2008), researchers have recently begun to investigate opinion spam as well. Wu et al., 2010; 2009, Yoo and Gretzel). According to Jindal and Liu (2008), opinion spam is both widespread and distinct from e-mail and web spam. They train models using features based on the review text, reviewer, and product in the absence of gold-standard deceptive opinions to distinguish between duplicate opinions7, which are regarded as deceptive spam, and non-duplicate opinions, which are regarded as truthful. Wu and others Based on the distortion of popularity rankings, they (2010) propose a different method for detecting deceptive opinion spam in the absence of gold-standard data. Because we compare gold-standard deceptive and truthful opinions, both of these heuristic evaluation methods are unnecessary in our work.

Yoo and Gretzel (2009) manually compare the psychologically relevant linguistic differences between 40 honest and 42 false hotel reviews using a standard statistical test. For our automated deception classifiers, on the other hand, we create a much larger dataset of 800 opinions. Psycholinguistic deception detection, a related task, has also been the subject of research. Newman and others 2003), and then Mihalcea and Strapparava (2009), ask participants to share both their true and false perspectives on personal issues (like their position on the death penalty, for instance). Zhou and co. 2004; 2008) think about computer-mediated deception in role-playing games that are meant to be played over email and instant messaging. However, in addition to comparing n-gram–based deception classifiers to a random guess baseline of 50%, these studies also evaluate and compare the performance of human judges (described in Section 3.3) and two other computational approaches (described in Section 4). Last but not least, automatic methods for determining review quality have been directly studied (Weimer et al., 2007), and in situations where assistance is needed (Danescu-Niculescu-Mizil et al., 2009; Kim and co., 2006; O'Mahony and Smyth, 2009) and trustworthiness (Weerkamp and De Rijke, 2008). Unfortunately, the majority of quality measures used in those works are solely based on human judgments, which we find in Section 3 to be inadequately calibrated for detecting opinion spam that is misleading. According to E. D. Wahyuni (2016), the rapid expansion of the Internet has had an impact on many of our day-to-day activities. One of the rapidly growing industries is e-commerce. On most e-commerce websites, customers can write reviews of a service. You can use these reviews as a source of information. Businesses, for instance, can use it to decide how to design their products or services, and potential customers can use it to decide whether to purchase or use a product. Sadly, certain individuals misjudged the meaning of the survey and endeavored to undermine the item or increment its prevalence by composing counterfeit audits. Using a survey's text and rating property, this study intends to identify fake surveys for a product. In a nutshell, the proposed system (ICF++) will measure the honesty of a review, the trustworthiness of reviewers, and the dependability of a product. The authenticity of a review will be evaluated using the text mining and opinion mining techniques. The experiment shows that the proposed system is more accurate than the iterative computation framework (ICF) method.According to M. Crawford (2016), one of the most significant sources of consumer information on a variety of goods and services is fast becoming online reviews. Spammers and unethical business owners now have more opportunities to create fake reviews to artificially promote their own products and services or smear those of their rivals because of their increased importance. Numerous studies on the most effective machine learning algorithms for detecting review spam have been conducted in response to this growing issue. The transformation of reviews into word vectors, which has the potential to produce hundreds of thousands of features, is a feature that runs through the majority of these studies. However, little research has been done on how to reduce the size of the feature subset to a manageable level or how to do so in the best way. In the review spam domain, we examine two distinct approaches to reducing the size of feature subsets. Word-frequency based feature selection and filter-based feature rankers are two of the methods. We demonstrate that there is no one-size-fits-all strategy for selecting features, and the most effective method for reducing the size of the feature subset is determined by the classifier employed and the

desired size. Additionally, it was discovered that the feature selection method selected had a significant impact on the size of the feature subset.

According to M. Luca and G. Zervas (2016), consumer reviews now play a role in everyday decision-making. However, when businesses fabricate reviews for themselves or their rivals, the credibility of these reviews is fundamentally undermined. Using two complementary approaches and datasets, we investigate the financial incentives for review fraud on the well-known review platform Yelp. We start by looking at restaurant reviews that Yelp's filtering algorithm has determined to be suspicious or fake. We use these reviews as a proxy for review fraud (for which we provide evidence). We discuss four primary findings. First, Yelp filters about 16% of restaurant reviews. These reviews are more extreme than other reviews (favorable or unfavorable), and the number of suspicious reviews has increased significantly over time. Second, a restaurant is more likely to engage in review fraud if it has a poor reputation, such as a lack of reviews or recent negative reviews. Thirdly, review fraud is less likely to occur at chain restaurants because these establishments receive less support from Yelp. Fourth, restaurants are more likely to receive negative fake reviews when there is more competition. We look at businesses that were found by Yelp to be soliciting fake reviews using a separate dataset. These data provide additional insight into the economic factors that influence a company's decision to leave fake reviews and back up our main findings.

According to A. j. Minnich (2015), online reviews of goods and services can be extremely beneficial to customers, but they must be protected from manipulation. The majority of studies to date have concentrated on examining online reviews from a single hosting company. How could information from multiple review hosting sites be utilized? In our work, this is the main question. As a response, we create a methodical approach to merging, comparing, and evaluating reviews from various hosting sites. We use more than 15 million hotel reviews from more than 3.5 million people across three well-known travel websites. Our work has three main focuses: a) we introduce the True View score as proof of concept that cross-site analysis can better inform the end user; b) we conduct arguably the first extensive study of cross-site variations using real data; c) we develop a hotel identity-matching method with 93% accuracy; and d) we develop novel features capable of identifying cross-site discrepancies effectively. Our findings demonstrate: 1) When we use multiple sites instead of just the three, we find seven times as many suspicious hotels, and 2) 20% of the hotels that appear on all three sites appear to have a low trustworthiness score. Our work is an early attempt to investigate the benefits and drawbacks of using multiple reviewing sites for better decision-making.

According to R. Shebuti (2015), testimonials from "real" people can be found in online reviews, which can help other consumers make decisions. However, opinion spam has become a widespread issue as a result of the financial rewards that come with positive reviews. Often, paid spam reviewers write fake reviews to unfairly promote or downgrade particular businesses or products. Behavioral footprints, relational ties between agents in a review system, and linguistic clues of deception have all been successfully used in previous approaches to opinion spam. In this work, we propose a new holistic strategy called Spangle that uses clues from all metadata (text, timestamp, rating) and relational data (network) in a unified framework to identify spam-targeted products, suspicious users, and reviews. In addition, our approach is capable of clearly and seamlessly integrating semi-supervision—that is, a (minimal) set of labels if they are available—without the need for additional training or modifications to its underlying algorithm. On three real-world Yelp.com review datasets with filtered (spam) and recommended (non-spam) reviews, Spangle significantly outperforms several baselines and cutting-edge techniques, demonstrating its electiveness and scalability. This is, to the best of our knowledge, the largest-ever quantitative evaluation of the opinion spam issue.

According to B. Viswanath (2014), Facebook ads and liked posts are examples of crowd-sourced information that users increasingly rely on. As a result, there is now a market for black hat methods of promotion using compromised and fake accounts, such as Sybil, and collusion networks. The majority of the currently available methods for detecting such behavior rely on supervised or semi-supervised learning over known or hypothesized attacks. They are unable to recognize attacks that the operator missed while labeling or when the attacker alters their plan. To differentiate potentially bad behavior from normal behavior, we propose applying unsupervised anomaly detection methods to user behavior. A method based on Principal Component Analysis (PCA) that accurately models the behavior of normal users and flags significant deviations from it as anomalous is presented by us. We tentatively approve that typical client conduct (e.g., classifications of Facebook pages preferred by a client, pace of like action, and so forth.) is contained within a PCA-compatible low-dimensional subspace. Using a lot of real-world data from Facebook, we show that our method works and is practical: With no a priori labeling, we are able to detect a wide range of attacker tactics, including fake, compromised, and coordinated Facebook identities, with low false-positive rates. Finally, we apply our method to Facebook ad click-spam detection and discover that a surprising number of clicks come from unusual users.

Ch. According to Xu and J. Zhang (2014), spam campaigns discovered on popular product review websites like Amazon.com, where a group of online posters are hired to collaboratively craft deceptive reviews for some target products, have attracted increasing attention from both industry and academia. The objective is to influence the targets' perceived reputations in their favor. Pairwise features, which have the potential to capture the underlying correlations among colluders, are either ignored or simply not explored sufficiently in the literature. Numerous efforts have been made to identify these colluders by extracting point-wise features from individual reviewers or reviewer-groups. Due to their nature as correlated components of spam campaigns, we discovered that pairwise features can be more robust at modeling the relationships between conspirators. In light of some collusion signals that can be found in the rating behaviors and linguistic patterns of reviewers, we investigate multiple heterogeneous pairwise features in his paper. Additionally, these pairwise features can be utilized in an intuitive, unsupervised collusion detection framework that has been proposed. Extensive testing on a real dataset demonstrates our method's efficacy and satisfactory superiority over several rivals.

According to H. Li (2014), online reviews have become an increasingly significant resource for product design and decision-making. However, opinion spam frequently targets reviews systems. Although supervised learning has been used for years to study fake review detection, large-scale dataset ground truth is still unavailable, and the majority of current supervised learning methods are based on pseudo rather than real fake reviews. With filtered reviews from Damping's fake review detection system, we present the first reported work on fake review detection in Chinese in collaboration with Dianping1, the largest review hosting site in China. The recall of Damping's algorithm is difficult to determine, despite its

high precision. This indicates that while the unknown set of fake reviews may not all be genuine, all fake reviews detected by the system are almost certainly fake. It is more appropriate to treat the unknown set as an unlabeled set due to the possibility that it contains numerous fake reviews. This necessitates the PU learning model, which uses positive and unlabeled examples to teach. We first propose a collective classification algorithm known as Multi-typed Heterogeneous Collective Classification (MHCC) and then extend it to Collective Positive and Unlabeled learning (CPU) by taking advantage of the intricate dependencies that exist between reviews, users, and IP addresses. Our experiments are based on actual customer feedback from 500 Shanghai, China, restaurants. In both PU and non-PU learning settings, the results demonstrate that the F1 scores of strong baselines can be significantly improved by our proposed models. It is simple to apply our models to other languages because they only make use of features that are independent of the language.

According to G. Fei (2013), user feedback is increasingly coming from online product reviews. Imposters have been writing deceptive or fake reviews to promote or downgrade certain target goods or services in order to make money or gain fame. Review spammers are such imposters. Several solutions to the issue have been proposed over the past few years. In this work, we employ a novel strategy that takes advantage of the hurried nature of reviews to identify review spammers. Either the sudden popularity of a product or spam attacks can cause waves of reviews. In the same way that genuine reviewers frequently appear alongside other genuine reviewers, spammers and reviews appearing in a burst are frequently associated. This makes it possible for us to establish a network of reviewers who appear in various bursts. The Loopy Belief Propagation (LBP) method is used to determine whether a reviewer is a spammer in the graph after modeling reviewers and their concurrence in bursts as a Markov Random Field (MRF). In addition, we use the LBP framework for network inference to propose a number of features and make use of feature induced message passing. In addition, we propose a novel evaluation strategy that uses supervised review classification to automatically evaluate the identified spammers. In addition, we employ domain experts to conduct an in-person assessment of the identified spammers and non-spammers. The proposed method outperforms strong baselines, as shown by the classification and human evaluation results, indicating its effectiveness.

According to M. Ott (2012), user-generated online reviews have a growing impact on consumers' purchasing decisions. As a consequence of this, there has been an increase in concerns regarding the potential for deceptive opinion spam|citations reviews to be posted, which are reviews that have been intentionally written to sound authentic in order to deceive the reader. However, despite the fact that this practice has received a lot of public attention and concern, very little is known about the actual rate of deception in online review communities and the factors that contribute to it. We use a deception classifier and a generative deception model that we propose to investigate the prevalence of deception in six well-known online review communities: Priceline, Expedia, Hotels.com, Orbitz, TripAdvisor, and Yelp. Based on economic signaling theory, we also propose a theoretical model of online reviews in which consumer reviews act as a signal to a product's true, unknown quality, reducing the inherent information asymmetry between consumers and producers. We find that deceptive opinion spam is becoming a bigger problem all over, but the rate of growth varies from community to community. We argue that the different signaling costs of deception, such as posting requirements, for each review community drive these rates. The prevalence of deception effectively decreases when measures are taken to increase the cost of signaling, such as filtering reviews written by first-time reviewers.

3. TECHNIQUE AND ALGORITHMS

- **Preprocessing**

- The tweets that are imported from the Twitter API into the database by this algorithm contain unnecessary words, whitespace, hyperlinks, and unique characters. We must first complete the separation process by removing all unnecessary words, whitespace, hyperlinks, and unusual characters.
- The goal of the preprocessing steps is to get the feature extraction and word bagging from the samples started. The reduction of the final number of features extracted is one of the primary goals. Indeed, feature reduction is essential for topic modeling and sentiment analysis predictions to be more accurate. Highlights are utilized to address the examples, and the more the calculation will be prepared for a particular element the more precise the outcomes will be. Therefore, it is convenient to combine two features that are similar into a single unique feature. Additionally, a feature can be removed from the list of words if it is irrelevant to the analysis.
- Letters in lowercase: Going through all of the data and changing each uppercase letter to its lowercase counterpart is the first step in the preprocessing process. The program will treat "data" and "Data" as distinct words when processing a word because the analysis is case-sensitive. It is essential to consider these two terms to share the same characteristics. In any other case, the algorithms will have an impact on emotions that may be distinct from these two words. Take, for instance, these three sentences: data are "good," "awesome," and "bad." Positive sentences use the word "data" in the first and second sentences, while negative sentences use the word "data" in the third. The algorithm will make an educated guess about whether sentences containing "data" are more likely to be positive or negative. The algorithm would have been able to guess that the fact that the sentence contains "data" is not very important for determining whether or not the sentence is positive if the uppercase letters had been removed. Due to the fact that the data were retrieved from Twitter, this preprocessing step is even more crucial. Web-based entertainment clients are much of the time writing in capitalized regardless of whether it isn't needed, consequently this preprocessing step will betterly affect virtual entertainment information than other "traditional" information.
- Remove user references and URLs: Users can include URLs, hashtags, and user references in their messages on Twitter. When analyzing a text's content, user references and URLs are typically irrelevant. As a result, this preprocessing step uses regular expressions to locate and substitute "URL" for each URL and "AT_USER" for each user reference, thereby reducing the total number of features extracted from the corpus [2]. Since the "#" characters will be removed during the tokenization process, the hashtags are not removed because they frequently contain a word that is pertinent to the analysis. Eliminate digits: It is possible to eliminate digits from the sentences because they are irrelevant to data analysis. Additionally, removing digits from words may enable the algorithm to associate two features that would otherwise have been considered distinct. For instance, some data may include "iphone" while others may include "iphone7." The tokenization procedure, which will be described in more detail later.
- Get rid of stop words: Stop words are frequently removed from the sample during natural language processing. These stop words are words that are frequently used in a language and are not relevant for topic modeling and

sentiment analysis, two natural language processing techniques [10]. The number of features extracted from the samples can be reduced by removing these words.

- Self-Learning and Word Standardization System In this algorithm, the word reference (first emphasis dictionary) must be implemented first. In the lexicon, the majority of the positive, negative, and things must be introduced. Without prepared information (introduction of words), all major datum and information mining endeavors fail, so the inclusion of prepared information is essential. In the self-learning framework, we institutionalize words; in this case, we only consider the word rather than its past, present, or future status.

Sentiment Analysis: Pre-processed tweets are taken one at a time from the database in this algorithm. First, we need to check each word individually to see if it is a thing or not. If it is, we will remove it from the specific review. The remainder of the words that were examined with assessment writing followed, regardless of whether those words refer to a particular opinion, a negative conclusion, or an impartial feeling. Due to the greater number of positive, negative, and impartial checks, the rest of the words in the tweet that have nothing to do with any of the hypotheses will be relegated to a transitory conclusion. In the second cycle, if a word crosses the line between positive, negative, or neutral, it stays in the lexicon forever as a development.

Calculating the cosine similarity The cosine similarity is a measure of the cosine of the angle between two non-zero vectors in an inner product space. Any angle in the range of [0,] radians has a cosine greater than or equal to 1. Therefore, it is a measurement of orientation rather than magnitude: There is a cosine similarity of 1 between two vectors oriented in the same direction, 0 between two vectors oriented at 90 degrees, and -1 between two vectors oriented diametrically opposed, regardless of their magnitude. In positive space, where the result is neatly bounded, the cosine similarity is especially useful. The term "direction cosine" is the source of the name: When unit vectors are parallel, they are maximally "similar" and when they are orthogonal (perpendicular), they are maximally "dissimilar." This is comparable to the cosine, which has a maximum value of unity and is uncorrelated when the segments are perpendicular to one another.

The cosine similarity is most frequently utilized in high-dimensional positive spaces, where these bounds are valid for any number of dimensions. For instance, in text mining and information retrieval, each term is notionally given a distinct dimension, and a document is represented by a vector whose value in each dimension is proportional to the number of times the term appears in the document. The measure of cosine similarity then provides a useful indication of how likely it is that two documents will share the same subject matter. In the field of data mining, the method is also used to measure cluster cohesion.

Algorithm Step in Sentiment Analysis

Step 1: Get some sentiment examples
As for every supervised learning problem, the algorithm needs to be trained from labeled examples in order to generalize to new data.

Step 2: Extract features from examples
Transform each example into a feature vector. The simplest way to do it is to have a vector where each dimension represents the frequency of a given word in the document.

Step3: Train the parameters
This is where your model will learn from the data. There are multiple ways of using features to generate an output, but one of the simplest algorithms is logistic regression. Other well-known algorithms are Naive Bayes. In the simplest form, each feature will be associated with a weight. Let's say the word "love" has a weight equal to +4, "hate" is -10, "the" is 0 ... For a given example, the weights corresponding to the features will be summed, and it will be considered "positive" if the total is > 0, "negative" otherwise. Our model will then try to find the optimal set of weights to maximize the number of examples in our data that are predicted correctly.
If you have more than 2 output classes, for example if you want to classify between "positive", "neutral" and "negative", each feature will have as many weights as there are classes, and the class with the highest weighted feature sum wins.

Step 4: Test the model
After we have trained the parameters to fit the training data, we have to make sure our model generalizes to new data, because it's really easy to over fit. The general way of regularizing the model is to prevent parameters from having extreme values.

Algorithm Step in Cosine Similarity

Step 1: Data Preparation
As with the k-means section, we will limit the number of attributes in the data set to A3 and A4 (petal length and petal width) using the Select Attribute operator, so that we can visualize the cluster and better understand the clustering process.
Step 2: Clustering Operator and Parameters
The modeling operator is available in the Modeling > Clustering and Segmentation folder, and is labeled DBSCAN. The allowing parameters can be configured in the model operator:

- Epsilon (ε): Size of the high-density neighborhood. The default value is 1.
- MinPoints: Minimum number of data objects within the epsilon neighborhood to qualify as a cluster.
- **Distance measure:** The proximity measure can be specified in this parameter. The default and most common measurement is Euclidean distance. Other options here are Manhattan distance, Jaccard coefficient, and cosine similarity for document data.
- **Add cluster as attributes:** To append cluster labels into the original data set. Turing on this option is recommended for later analysis.

Step 3: Evaluation (Optimal)
Similar to k-means clustering implementation, we can evaluate the effectiveness of clustering groups using average within cluster distance. In Rapid Miner, the Cluster Density Performance operator under Evaluation > Clustering is available for performance evaluation of cluster groups generated by Density algorithms. The clustering model and labeled data set is connected to performance operator for cluster evaluation. Additionally, to aid the calculation, performance operator expects Similarity Measure object. A similarity measure vector is a distance measure of every example data object with the other data object. The similarity measure can be calculated by using Data to Similarity Operator on the example data set.

Step 4: Execution and Interpretation

After the outputs from the performance operator have been connected to the result ports, the model can be executed.

6. CONCLUSION

Based on the results of a noisy deception classifier, we have provided a general framework for estimating the prevalence of deception in online review communities in this review. We provided the first empirical study of the magnitude and underlying factors of deceptive opinion spam by examining the prevalence of deception among positive reviews in six popular online review communities using this framework. Based on economic signaling theory, we have also proposed a theoretical model of online reviews as a sign of a product's true (unknown) quality. We have specifically defined the signal cost of positive online reviews as a function of the costs of posting and the benefits of exposure to the review community. We have proposed two additional hypotheses based on this theory, both of which are supported by our findings. We first find that review communities with low signal costs (low posting requirements, high exposure) are more likely to deceive than communities with higher signal costs. Second, we discover that we can effectively reduce both the prevalence and growth rate of deception in a review community by increasing the signal cost of that community, such as by excluding reviews written by first- or second-time reviewers.

REFERENCE

- M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa, (2016). Reducing Feature set Explosion to Faciliate Real-World Review Sapm Detection. In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference.
- M. Luca and G. Zervas, (2016). Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud., SSRN Electronic Journal.
- M. Ott, C. Cardie, and J. T. Hancock, (2012). Estimating the prevalence of deception in online review communities. In ACM WWW, 2012.
- M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, (2011). Finding deceptive opinion spam by any stretch of the imagination.In ACL.
- M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Montesi, (2015). Spreading processes in multilayer networks. In IEEE Transactions on Network Science and Engineering. 2(2):65–83.
- N. Jindal and B. Liu. Opinion Spam and Analysis, (2008). In WSDM.
- N. Jindal, B. Liu, and E.-P. Lim, (2012). Finding unusual review patterns using unexpected rules. In ACM CIKM.
- R. Hassanzadeh, (2014). Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic. Queensland University of Technology, Nov.
- R. Shebuti and L. Akoglu, (2015). Collective opinion spam detection: bridging review networksand metadata. In ACM KDD.
- S. Feng, L. Xing, A. Gogar, and Y. Choi, (2012). Distributional footprints of deceptive product reviews. In ICWSM.
- S. Feng, R. Banerjee and Y. Choi, (2012). Syntactic stylometry for deception detection. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers; ACL.
- S. Mukherjee, S. Dutta, and G. Weikum, (2016). Credible Review Detection with Limited Information using Consistency Features, In book: Machine Learning and Knowledge Discovery in Databases.
- S. Xie, G. Wang, S. Lin, and P. S. Yu, (2012). Review spam detection via temporal pattern discovery. In ACM KDD.
- Y. Sun and J. Han, (2012). Mining Heterogeneous Information Networks; Principles and Methodologies, In ICCCE.
- Y. Sun and J. Han, (2009). Rankclus: integrating clustering with ranking for heterogeneous information network analysis. In Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology.
- Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, (2011). Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. In VLDB.
- A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, (2014). Detection of review spam: A survey. Expert Systems with Applicants, Elsevier.
- A.j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, (2015), Trueview: Harnessing the power of multiple review sites. In ACM WWW.
- A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh (2013), Spotting opinion spammers using behavioral footprints. In ACM KDD.
- A. Mukherjee, B. Liu, and N. Glance, (2012), Spotting Fake Reviewer Groups in Consumer Reviews. In ACM WWW, 2012.
- A. Mukerjee, V. Venkataraman, B. Liu, and N. Glance, (2013), What Yelp Fake Review Filter Might Be Doing?, In ICWSM, 2013.
- B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, (2014). Towards detecting anomalous user behavior in online social networks. In USENIX.
- C. L. Lai, K. Q. Xu, R. Lau, Y. Li, and L. Jing, (2011). Toward a Language Modeling Approach for Consumer Review Spam Detection. In Proceedings of the 7th international conference on e-Business Engineering.
- C. Luo, R. Guan, Z. Wang, and C. Lin, (2014). HetPathMine: A Novel Transductive Classification Algorithm on Heterogeneous Information Networks. In ECIR.
- Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pairwise features, (2014). In SIAM International Conference on Data Mining.
- E. D. Wahyuni and A. Djunaidy, (2016). Fake Review Detection From a ProductReview Using Modified Method of Iterative Computation Framework. In Proceeding MATEC Web of Conferences.
- E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, (2010). Detecting product review spammers using rating behaviors. In ACM CIKM.
- F. Li, M. Huang, Y. Yang, and X. Zhu, (2011). Learning to identify review spam. Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI.
- G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, (2013). Exploiting burstiness in reviews for review spammer detection. In ICWSM.
- G. Wang, S. Xie, B. Liu, and P. S. Yu, (2011). Review graph based online store review spammer detection. IEEE ICDM.
- H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, (2014). Spotting fake reviews via collective PU learning. In ICDM.
- H. Xue, F. Li, H. Seo, and R. Pluretti, (2015). Trust-Aware Review Spam Detection. IEEE Trustcom/ISPA .

- J. Donfro, (2015). A whopping 20 % of yelp reviews are fake. http://www.businessinsider.com/20-percent-of-yelp-reviews-fake-2013-9. Accessed: 2015-07-30.
- K. Weise. A Lie Detector Test for Online Reviewers, (2016). http://bloom.bg/1KAxzhK. Accessed: 2016-12-16.
- L. Akoglu, R. Chandy, and C. Faloutsos, (2013). Opinion fraud detection in online reviews bynetwork effects. In ICWSM.
- M. Crawford, T. D. Khoshgoftar, J. N. Prusa, A. Al. Ritcher, and H. Najada, (2015). Survey of Review Spam Detection Using Machine Learning Techniques. Journal of Big Data.

.