INTERNATIONAL JOURNAL OF
TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

Detecting Spammer Groups From Product Reviews: A Partially Supervised Learning Model

BALLA H S NAGA VARA PRASAD
PG Scholar, Department of M.C.A,
S.K.B.R P.G College,
Amalapuram, E.G.Dt., A.P, India.
E-Mail: bhsnvprasad2001@gmail.com

Mr. NAGA. SRINIVASA RAO*
Asst. Professor, Dept of M.C.A,
S.K.B.R P.G College,
Amalapuram, E.G.Dt., A.P, India.
E-Mail:naagaasrinu@gmail.com

Abstract

—Today, online product reviews play a crucial role in consumers' purchasing decisions. A high proportion of positive reviews leads to significant sales growth, while negative reviews lead to lost sales. Driven by the immense financial gains, many spammers try to promote their products or discredit their competitors' products by posting fake and biased online reviews. By registering multiple accounts or releasing tasks on crowd-sourcing platforms, many individual spammers could organize as spammer groups to collectively manipulate product reviews and cause more damage. Existing work on spammer group detection extracts candidate spammer groups from review data and identifies the real spammer groups using unsupervised spamity ranking methods. In fact, according to previous research, it's easier to tag a small number of spammer groups than one might think, yet few methods attempt to make good use of this important tagged data.This project proposes a partially supervised learning model (PSGD) to identify spammer groups. By marking some spammer groups as positive instances, PSGD applies positive unmarked learning (PU learning) to generate a spammer group detector classified positive instances (marked spammer groups) and unmarked instances (unmarked groups) detector. to investigate. In particular, extract reliable negative sentences related to the positive cases and the characteristic features. By combining the positive instances, extracted negative instances, and unlabeled instances, convert the PU learning problem to the well-known semi-supervised learning problem, and then use a naive Bayesian model and an EM algorithm to train a classifier for spammer group detection. Experiments with real Amazon.cn data sets show that the proposed PSGD is effective and outperforms state-of-the-art spammer group detection methods

Index Terms—**Social network analysis, spammer detection, spambot detection, social network security.**

I. Introduction

In e-commerce platforms, online product reviews becomemore and more important as the purchase decisions of the customers are strongly influenced by these reviews. Due to the financial incentives, many imposters try to game thesystems and consumers by posting biased ratings and reviewsto promote their products or demote their competitors' products. These imposters, also called *Review Spammers* or*Opinion Spammers*, become more and more damager as theycould be organized by crowdsourcing tasks. As there aremany accounts, the organized spammers, called *SpammerGroup*, could take total control of the sentiment on their targetproducts with little abnormal behavior.Although many efforts have been done for review spamand individual spammer detection, limited attentionhas been received at the spammer group detection.Generally, as there are usually no labeled instances (groups),most existing work find spammer group candidates first,and then use unsupervised ranking methods to identify realspammer groups from these candidates. Easily label some groupsmanually to obtain some labeled instances (i.e., labeled spammergroups or non-spam groups). It is obvious that combiningthese labeled instances and other unlabeled groupswill significantly improve the accuracy of spammer group detection.Simultaneously utilizing labeled and unlabeled data isa typical problem of partially supervised learning. Strictlyspeaking, there are two types of partially supervised learningaccording to the constitution of labeled data. Onetype is that the labeled data contains the instances of allclasses (e.g., containing both spammer and non-spammergroups in this paper), which is commonly known as semi-supervisedlearning. The second type is that the labeleddata only contains positive instances (e.g., spammer groups) and need to learn from the positive and unlabeled instances. This project calls the second type of partiallysupervised learning as Positive Unlabeled Learning(PU-Learning for short, where P and U stand for positiveand unlabeled instances, respectively). Since labelingspammer groups is much easier than labeling non-spammer groups, it adopt PU-Learning as the main technique todetect spammer groups without labeling any non-spammergroups.This project proposes a Partially Supervised learningbased Spammer Group Detection (PSGD) model. Like mostexisting spammer group detection methods, use frequentitem mining (FIM) to extract spammer group candidates,and then apply PU-Learning to detect real spammer groupsfrom these candidates. Specifically, manually label somespammer groups from the found group candidates as positiveinstances. Then, supervised by these positive instances, design an algorithm to automatically extract reliable negativeset (*RN*) which consists of only non-spammer groups. Combinethe positive and negative instances, the PU-Learningproblem could be converted into the well-known semisupervisedlearning problem, then the Naive Bayesian modeland Expectation Maximization (EM) algorithm are used to train a classifier to detect spammer groups.

II. RELATED WORKS

Impact of online consumer reviews on sales

F. Zhu and X. Zhang [1] ‘‘Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics,” in this work how product and consumer characteristics moderate the influence of online consumer reviews on product sales using data from the video game industry. The findings indicate that online reviews are more influential for less popular games and games whose players have greater Internet experience. The differential impact of consumer reviews across products in the same product category and suggests that firms' online marketing strategies should be contingent on product and consumer characteristics.

Temporal dynamics of opinion spamming

K. C. Santosh and A. Mukherjee [2] ''on the temporal dynamics of opinion spamming: Case studies on yelp,'' In this work recently, the problem of opinion spam has been widespread and has attracted a lot of research attention. While the problem has been approached on a variety of dimensions, the temporal dynamics in which opinion spamming operates is unclear.

How does buffered spamming operate for entities that need spamming to retain threshold popularity and reduced spamming for entities making better success? They analyze these questions in the light of time-series analysis on Yelp.

This work analyses discover various temporal patterns and their relationships with the rate at which fake reviews are posted. Building on analyses, they employ vector auto regression to predict the rate of deception across different spamming policies. They explore the effect of filtered reviews on (long-term and imminent) future rating and popularity prediction of entities. The results discover novel temporal dynamics of spamming which are intuitive, arguable and also render confidence on Yelp's filtering. Lastly, leverage discovered temporal patterns in deception detection. Experimental results on large-scale reviews show the effectiveness of this approach that significantly improves the existing approaches

Opinion spam and analysis

N. Jindal and B. Liu [3], ''Opinion spam and analysis,'' in this work past few years, sentiment analysis and opinion mining becomes a popular and important task. These studies all assume that their opinion resources are real and trustful.

However, they may encounter the faked opinion or opinion spam problem. They studythis issue in the context of product review mining system. On product review site, people may write faked reviews, called review spam, to promote their products, or defame their competitors' products. It is important to identify and filter out the review spam.

Previous work only focuses on some heuristic rules, such as helpfulness voting, or rating deviation, which limits the performance of this task. They exploit machine learning methods to identify review spam. Toward the end, manually build a spam collection from crawled reviews. First analyze the effect of various features in spam identification. We also observe that the review spammer consistently writes spam.

This provides us another view to identify review spam: identify if the author of the review is spammer. Based on this observation, provide a two view. Semi-supervised method, co-training, to exploit the large amount of unlabeled data. The experiment results show that proposed method is effective. This work designed machine learning methods achieve significant improvements in comparison to the heuristic baselines.

Learning to identify review spam

F. Li, M. Huang, Y. Yang, and X. Zhu [4], ‘‘Learning to identify review spam,’ in these work online reviews plays a crucial role in today’s electronic commerce. It is desirable for a customer to read reviews of products or stores before making the decision of what or from where to buy. Due to the pervasive spam reviews, customers can be misled to buy low-quality products, while decent stores can be defamed by malicious reviews. Observe that, in reality, a great portion of the reviewers write only one review .These reviews are so enormous in number that they can almost determine a store’s rating and impression.

In existing methods did not examine this larger part of the reviews. To address this problem, observe that the normal re- viewers’ arrival pattern is stable and uncorrelated to their rating pattern temporally. In contrast, spam attacks are usually busty and either positively or negatively correlated to the rating. Thus, they propose to detect such attacks via unusually correlated temporal patterns. Identify and construct multidimensional time series based on aggregate statistics, in order to depict and mine such correlations.

In this way, the singleton review spam detection problem is mapped to a abnormally correlated pattern detection problem. Proposes a hierarchical algorithm to robustly detect the time windows where such attacks are likely to have happened. The algorithm also pinpoints such windows in different time resolutions to facilitate faster human inspection. Experimental results show that the proposed method is effective in detecting singleton review attacks. Discover that singleton review is a significant source of spam reviews and largely affects the ratings of online store.

Review spam detection via temporal pattern discovery

S. Xie, G. Wang, S. Lin, and P. S. Yu [5] ‘‘Review spam detection via temporal pattern discovery,’’ in this work used by individuals and organizations for their decision making. However, due to the reason of profit or fame, people try to game the system by opinion spamming (e.g., writing fake reviews) to promote or to demote some target products. In recent years, fake review detection has attracted significant attention from both the business and research communities.

However, due to the difficulty of human labeling needed for supervised learning and evaluation, the problem remains to be highly challenging. This work proposes a novel angle to the problem by modeling spam city as latent. An unsupervised model, called Author Spam city Model (ASM), is proposed. It works in the Bayesian setting, which facilitates modeling spam city of authors as latent and allows us to exploit various observed behavioral footprints of reviewers. The intuition is that opinion spammers have different behavioral distributions than non-spammers.

This creates a distributional divergence between the latent population distributions of two clusters: spammers and non-spammers. Model inference results in learning the population distributions of the two clusters. Several extensions of ASM are also considered leveraging from different priors. Experiments on a real-life Amazon review dataset demonstrate the effectiveness of the proposed models which significantly outperform the state-of-the-art competitors.

Spotting opinion spammers using behavioral footprints

A. Mukherjee et al [6] ‘‘Spotting opinion spammers using behavioral footprints,” in these work User-generated online reviews can play a significant role in the success of retail products, hotels, restaurants, etc.

However, review systems are often targeted by opinion spammers who seek to distort the perceived quality of a product by creating fraudulent reviews. Propose a fast and effective framework, fraud eagle, for spotting fraudsters and fake reviews in online review datasets. in this method has several advantages:

it exploits the network effect among reviewers and products, unlike the vast majority of existing methods that focus on review text or behavioral analysis, it consists of two complementary steps; scoring users and reviews for fraud detection, and grouping for visualization and sense making, it operates in a completely unsupervised fashion requiring no labeled data, while still incorporating side information if available, and it is scalable to large datasets as its run time grows linearly with network size. Demonstrate the effectiveness of this framework on synthetic and real datasets; where fraud eagle successfully reveal fraud-bots in a large online app review database.

Opinion fraud detection in online reviews by network effects

L. Akoglu, R. Chandy, and C. Faloutsos, [7] ‘‘Opinion fraud detection in online reviews by network effects,’’ big part of people rely on available content in social media in their decisions (e.g. reviews and feedback on a topic or product). The possibility that anybody can leave a review provides a golden opportunity for spammers to write spam reviews about products and services for different interests.

Identifying these spammers and the spam content is a hot topic of research and although a considerable number of studies have been done recently toward this end, but so far the methodologies put forth still barely detect spam reviews, and none of them show the importance of each extracted feature type. This study, proposes a novel framework, named Net Spam, which utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks. Using the importance of spam features help to obtain better results in terms of different metrics experimented on real-world review datasets from Yelp and Amazon websites.

The results show that Net Spam outperforms the existing methods and among four categories of features; including review-behavioral, user-behavioral, review linguistic, nuser-linguistic, the first type of features performs better than the other categories.

Net Spam: A network-based spam detection framework for reviews in online social media

S. Shehnepoor, M. Salehi, R. Farahbakhsh, and N. Crespi[8] ‘‘Net Spam: A network-based spam detection framework for reviews in online social media,’’ In this work Driven by profits, spam reviews for product promotion or suppression become increasingly rampant in online shopping platforms.

This work focuses on detecting hidden spam users based on product reviews. In the literature, there have been tremendous studies suggesting diversified methods for spammer detection, but whether these methods can be combined effectively for higher performance remains unclear. Along this line, a hybrid PU-learning-based Spammer Detection (hPSD) model is proposed in this work. On one hand, hPSD can detect multi-type spammers by injecting or recognizing only a small portion of positive samples, which meets particularly real-world application scenarios. More importantly, hPSD can leverage both user features and user relations to build a spammer classifier via a semi-supervised hybrid learning framework.

Experimental results on amazon data sets with shilling injection show that hPSD outperforms several state-of-the-art baseline methods. In particular, hPSD shows great potential in detecting hidden spammers as well as their underlying employers from a real life Amazon data set. These demonstrate the effectiveness and practical value of hPSD for real-life applications

Spammers detection from product reviews: A hybrid model

Z. Wu, Y. Wang, Y. Wang, J. Wu, J. Cao, and L. Zhang [9] ‘‘Spammers detection from product reviews: A hybrid model,’’in this work Today’s e-commerce is highly depended on increasingly growing online customers’ reviews posted in opinion sharing websites.

This fact, unfortunately, has tempted spammers to target opinion sharing web- sites in order to promote and demote products. To date, different types of opinion spam detection methods have been proposed in order to provide reliable resources for customers, manufacturers and re- searchers. However, supervised approaches suffer from imbalance data due to scarcity of spam reviews in datasets, rating deviation based filtering systems are easily cheated by smart spammers, and content based methods are very expensive and majority of them have not been tested on real data hitherto.

The aim of this work is to propose a robust review spam detection system wherein the rating deviation, content based factors and activeness of reviewers are employed efficiently. To overcome the aforementioned drawbacks, all these factors are synthetically investigated in suspicious time intervals captured from time series of reviews by a pattern recognition technique. The proposed method could be a great asset in online spam filtering systems and could be used in data mining and knowledge discovery tasks as a standalone system to purify product review datasets.

These systems can reap benefit from the method in terms of time efficiency and high accuracy. Empirical analyses on real dataset show that the proposed approach is able to successfully detect spam reviews. Comparison with two of the current common methods, indicates that this method is able to achieve higher detection accuracy (F-Score: 0.86) while removing the need for having specific fields of Meta data and reducing heavy computation required for investigate purposes.

Detection of fake opinions using time series

A. Heydari, M. Tavakoli, and N. Salim [10] ‘‘Detection of fake opinions using time series, “in this work Online reviews play a crucial role in helping consumers evaluate and compare products and services.

This critical importance of reviews also incentivizes fraudsters (or spammers) to write fake or spam reviews to secretly promote or demote some target products and services. Existing approaches to detecting spam reviews and reviewers employed review contents, reviewer behaviors, star rating patterns, and reviewer-product networks for detection. This research, further discovered that reviewers' posting rates (number of reviews written in a period of time) also follow an interesting distribution pattern, which has not been reported before. That is, their posting rates are bimodal.

Multiple spammers also tend to collectively and actively post reviews to the same set of products within a short time frame, which call co-bursting. Furthermore, found some other interesting patterns in individual reviewers' temporal dynamics and their co-bursting behaviors with other reviewers. Inspired by these findings, first propose a two-mode Labeled Hidden Markov Model to model spamming using only individual reviewers' review posting times. Then extend it to the Coupled Hidden Markov Model to capture both reviewer posting behaviors and co-bursting signals.

In these experiments show that the proposed model significantly outperforms state-of-the-art baselines in identifying individual spammers. Furthermore, propose a co bursting network based on co-bursting relations, which helps detect groups of spammers more effectively than existing approaches.

Bimodal distribution and co-bursting in review spam detection

H. Li et al [11]., ‘‘Bimodal distribution and co-bursting in review spam detection,’’ in this work As the rapid development of China's e-commerce in recent years and the underlying evolution of adversarial spamming tactics, more sophisticated spamming activities may carry out in Chinese review websites. Empirical analysis, on recently crawled product reviews from a popular Chinese e-commerce website, reveals the failure of many state-of the- art spam indicators on detecting collusive spammers.

Two novel methods are then proposed: a KNN-based method that considers the pairwise similarity of two reviewers based on their group-level relational information and selects *k* most similar reviewers for voting; a more general graph-based classification method that jointly classifies a set of reviewers based on their pairwise transaction correlations. Experimental results show that both this methods promisingly outperform the indicator-only classifiers in various settings

Spotting fake reviewer groups in consumer reviews

A. Mukherjee, B. Liu, and N. Glance [12], ‘‘Spotting fake reviewer groups in consumer reviews,’’ in this work Opinionated social media such as product reviews are now widely used by individuals and organizations for their decision making.

However, due to the reason of profit or fame, people try to game the system by opinion spamming (e.g., writing fake reviews) to promote or demote some target products. For reviews to reflect genuine user experiences and opinions, such spam reviews should be detected. Prior works on opinion spam focused on detecting fake reviews and individual fake reviewers.

However, a fake reviewer group (a group of reviewers who work collaboratively to write fake reviews) is even more damaging as they can take total control of the sentiment on the target product due to its size.

This work studies spam detection in the collaborative setting, i.e., to discover fake reviewer groups. The proposed method first uses a frequent item set mining method to find a set of candidate groups. It then uses several behavioral models derived from the collusion phenomenon among fake reviewers and relation models based on the relationships among groups, individual reviewers, and products they reviewed to detect fake reviewer groups. Additionally built a labeled dataset of fake reviewer groups. Although labeling individual fake reviews and reviewers is very hard, to this surprise labeling fake reviewer groups is much easier.

Note that the proposed technique departs from the traditional supervised learning approach for spam detection because of the inherent nature of the problem which makes the classic supervised learning approach less effective. Experimental results show that the proposed method outperforms multiple strong baselines including the state of supervised classification, regression, and learning to rank algorithms.

Uncovering collusive spammers in Chinese review websites

C. Xu, et.al [13] ‘‘Uncovering collusive spammers in Chinese review websites, ‘In this work Community Question Answering (CQA) portals provide rich sources of information on a variety of topics.

However, the authenticity and quality of questions and answers (Q&as) has proven hard to control. In a troubling direction, the widespread growth of crowd sourcing websites has created a large-scale, potentially difficult-to-detect workforce to manipulate malicious contents in CQA. The crowd workers who join the same crowd sourcing task about promotion campaigns in CQA collusively manipulate deceptive Q&As for promoting a target (product or service). The collusive spamming group can fully control the sentiment of the target.

How to utilize the structure and the attributes for detecting manipulated Q&As? How to detect the collusive group and leverage the group information for the detection task? To shed light on these research questions, propose a unified framework to tackle the challenge of detecting collusive spamming activities of CQA.

First, interpret the questions and answers in CQA as two independent networks. Second, detect collusive question groups and answer groups from these two networks respectively by measuring the similarity of the contents posted within a short duration. Second using attributes (individual level and group-level) and correlations (user-based and content based), proposed a combined factor graph model to detect deceptive Q&As simultaneously by combining two independent factor graphs.

Discovering shilling groups in a real e-commerce platform

Y. Wang, Z. Wu, Z. Bu, J. Cao, and D. Yang [14] ‘‘Discovering shilling groups in a real e-commerce platform, ‘in this work As the use of recommender systems becomes generalized in society, the interest in varying the orientation of their recommendations is increasing.

There are shilling attacks' strategies that introduce malicious process in collaborative altering recommender systems in order to promote the own products or services or to discredit those of the competition. Academic research against shilling attacks has been focused in statistical approaches to detect the unusual patterns in user ratings.

Nowadays, there is a growing research area focused on the design of robust machine learning methods to neutralize the malicious process inserted into the system. This work proposes an innovative robust method, based on matrix factorization, to neutralize the shilling attacks. This method obtains the reliability value associated with each prediction of a user to an item. By monitoring the unusual reliability variations in the items prediction, can avoid promoting the shilling predictions to the erroneous recommendations

Partially supervised learning

B. Liu and W. S. Lee [15], ‘‘Partially supervised learning, ‘in this work investigate the following problem: Given a set of documents of a particular topic or class , and a large set _ of mixed documents that contains documents from class and other types of documents, identify the documents from class.

The key feature of this problem is that there is no labeled non document, which makes traditional machine learning techniques inapplicable, as they all need labeled documents of both classes.

Call this problem partiallysupervised classification. This work, show that this problem can be posed as a constrained optimization problem and that under appropriate conditions; solutions to the constrained optimization problem will give good solutions to the partially supervised classification problem.

III. Existing System

Existing methods find duplicateor near duplicate reviews to detect the review spam, wherethe similarity of reviews are mainly calculated by n-grambased review content comparison or probabilistic language model.

The review spam detection is deemed as a binary classification or ranking problem.Many content features and metadata of reviews, such asparts-of-speech (POS), term frequency and n-gram features, are used for classification or ranking.

Spammerdetection has the similar principle withreview spam detection, however, the features for classificationor ranking are mainly constructed from user behaviorsuch as review/rating posting time, rating deviation, burstreview ratio, reviewer burstiness and ratio of verified purchase(only in Amazon).

Existing work usually uses frequent itemset mining(FIM) to discover group candidates first, and then identifiesthe candidates as spammer or non-spammer groupsusing unsupervised ranking methods.

ISSN: 2057-5688

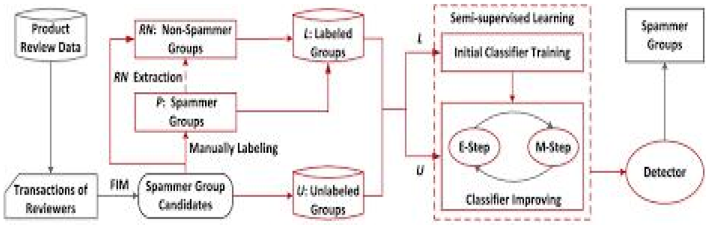
Labelingspammer groups is much easier than labeling individual spammers, learning a classifier from these positiveinstances and other unlabeled instances is a straightforwardway to improve the accuracy of spammer group detection.

IV. Proposed System

This project propose a partially supervised learningmodel to detect spammer groups, in which PU-Learning isapplied to extract reliable negative set and train a semi-supervised learning classifier as detector based on Naïve Bayesian model and EM algorithm. The detection of spammergroups usually contains two phases: (i) Discover spammergroup candidates, (ii) Identify the real spammer groups fromthe candidates.

Inthe context of spammer group detection, the reviewers areseen as the items and the reviewers who have co-revieweda particular product are regarded as a transaction. By mining frequent itemsets, find groups of reviewers whohave co-reviewed multiple products together as spammergroup candidates. Among the extracted candidates, somespammer groups are manually labeled to construct positiveinstances set, denoted as *P*. Then, by automatically extractingsome groups whose features are significantly different withinstances in *P*, the reliable negative set (denoted as *RN*)consisting of only non-spammer groups will be constructed.Combine *P* and *RN*, we will obtain a labeled data set (denotedas *L*) containing both positive and negative instances, andthe remainder spammer group candidates with unknownclasses will construct an unlabeled data set (denoted as *U*).Based on *L* and *U*, a semi-supervised learning classifieris trained to identify real spammer groups, which initials aNaive Bayes classifier on *L* and improves it on *U* using anExpectation Maximization (EM) algorithm.

System Architecture



V. MODULES

Preprocessing

In this module the given review data set is preprocessed. The data set contains the following attributes: Summary of the review, reviewerID, overall ratings, asin (product id), unixReviewTime, reviewText and reviewTime. This project only needs reviewer ID, product ID, review text, time and rating. After extracting this information, a transaction is build based on the reviewer ID and Product ID.

FIM

In this module, Frequent Itemset mining is applied to find out group candidates. Thequality and quantity of candidate groups are heavily dependent on the supportthreshold. If the threshold is much higher, there are less yet more suspiciouscandidates; otherwise, the candidates might contain a great deal of normal users. FIM take reviewer identifiers as items, and products as transactions. By setting the minimum support count, they can find candidate groups which have at least two reviewers and each reviewer at least reviews three common products.

Feature Extraction

This module extracts the following features from the spammer candidate group. They are Group Time Window (GTW), Group Deviation (GD), Group Early Time Frame (GETF), Group Size Ratio (GSR), Group Size (GS), Group Support Count(GSUP), Product Tightness (PT), Rating Variance (RV), Product Reviewer Ratio (PRR) and Average Active Interval (AAI).

Spammer Detection

This module detects the spammer group based on the extracted features. The classification algorithm Naïve Bayes, Multi-Layer Perceptron, Logistic Regression and Sequential Mining Optimization is used to analyze the performance of the proposed work.

VI. PROCESS MODEL

A spammer group consists of a set of reviewers who co-reviews a set of common products. Thus, the data mining technique Frequent Itemset Mining (FIM) could be utilized to extract the groups . However, since many users maybecoincidentallygroupedbecauseofthe similar interest, the groups extracted by FIM are only the spammer group candidates and need to be further checked to identify the real spammer groups. Therefore, the detection of spammer groups usually contains two phases: (i) Discover spammer group candidates, (ii) Identify the real spammer groups from the candidates. Our proposed PSGD model is also along this line. SMOTE to balance the dataset. In the context of spammer group detection, the reviewers are seen as the items and the reviewers who have co-reviewed a particular product are regarded as a transaction. By mining frequent itemsets, we find groups of reviewers who have co-reviewed multiple products together as spammer group candidates. Among the extracted candidates, some spammer groups are manually labeled to construct positive instancesset,denotedasP.Then,byautomatically extracting some groups whose features are significantly different with instances in P, the reliable negative set (denoted as RN) consisting of only non-spammer groups will be constructed. CombinePandRN,wewillobtainalabeleddataset (denoted as L) containing both positive and negative instances, andthe remainder spammer group candidates with unknown classes will construct an unlabeled data set (denoted as U). Based on L and U, a semi-supervised learning classifier is trained to identify real spammer groups, which initials a Naive Bayes classifier on L and improves it on U using an Expectation Maximization (EM) algorithm.

VII. CONCLUSION

This paper proposes a partially supervised learning basedmodel PSGD to detect spammer groups from productreviews. First, the PSGD model uses frequent item mining(FIM) to discover spammer group candidates from the reviewdata. Then, by manually labeling some spammer groups aspositive instances, the PSGD employs PU-Learning to construct a classifier from the positive and unlabeled instancesto identify the real spammer groups from group candidates.In particular, the PSGD

ISSN: 2057-5688

defines a feature strength function tomeasure the discriminative power of group features, and theniteratively removes instances containing high discriminativefeatures from the unlabeled instances set to obtain a reliablenegative set consisting of only non-spammer groups. By combiningthe positive, negative and unlabeled instances, weconvert the PU-Learning problem into the well-known semi-supervisedlearning problem, and employ Naive Bayesianmodel and EM algorithm to construct a classifier as spammergroup detector. Experiments on Amazon.cn demonstrate thatthe proposed PSGD model outperforms both supervised andunsupervised learning methods on spammer group detection.

VIII. REFERENCES

1. Improving product marketing by predicting early reviewers on E-Commerce websites
S. Kodati, M. Dhasaratham, V. V. S. S. Srikanth, and K. M. Reddy, "Improving product marketing by predicting early reviewers on E-Commerce websites," Deleted Journal, no. 43, pp. 17–25, Apr. 2024, doi: 10.55529/ijrise.43.17.25.

2. Kodati, Dr Sarangam, et al. "Classification of SARS Cov-2 and Non-SARS Cov-2 Pneumonia Using CNN." Journal of Prevention, Diagnosis and Management of Human Diseases (JPDMHD) 2799-1202, vol. 3, no. 06, 23 Nov. 2023, pp. 32–40, journal.hmjournals.com/index.php/JPDMHD/article/view/3406/2798, https://doi.org/10.55529/jpdmhd.36.32.40. Accessed 2 May 2024.

3. V. Srikanth, "CHRONIC KIDNEY DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS," IJTE, pp. 106–109, Jan. 2023, [Online]. Available: http://ijte.uk/archive/2023/CHRONIC-KIDNEY-DISEASE-PREDICTION-USING-MACHINE-LEARNING-ALGORITHMS.pdf

4. V. SRIKANTH, "DETECTION OF PLAGIARISM USING ARTIFICIAL NEURAL NETWORKS," International Journal of Technology and Engineering, vol. XV, no. I, pp. 201–204, Feb. 2023, [Online]. Available: http://ijte.uk/archive/2023/DETECTION-OF-PLAGIARISM-USING-ARTIFICIAL-NEURAL-NETWORKS.pdf

5. V. SRIKANTH, "A REVIEW ON MODELING AND PREDICTING OF CYBER HACKING BREACHES," IJTE, vol. XV, no. I, pp. 300–302, Mar. 2023, [Online]. Available: http://ijte.uk/archive/2023/A-REVIEW-ON-MODELING-AND-PREDICTING-OF-CYBER-HACKING-BREACHES.pdf

6. S. Kodati, M. Dhasaratham, V. V. S. S. Srikanth, and K. M. Reddy, "Detection of fake currency using machine learning models," Deleted Journal, no. 41, pp. 31–38, Dec. 2023, doi: 10.55529/ijrise.41.31.38.

7. "Cyberspace and the Law: Cyber Security." IOK STORE, iokstore.inkofknowledge.com/product-page/cyberspace-and-the-law. Accessed 2 May 2024.

8. “Data Structures Laboratory Manual.” IOK STORE, www.iokstore.inkofknowledge.com/product-page/data-structures-laboratory-manual. Accessed 2 May 2024.

9. Data Analytics Using R Programming Lab.” IOK STORE, www.iokstore.inkofknowledge.com/product-page/data-analytics-using-r-programming-lab. Accessed 2 May 2024.

10. V. Srikanth, Dr. I. Reddy, and Department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, 501301, India, “WIRELESS SECURITY PROTOCOLS (WEP,WPA,WPA2 & WPA3),” journal-article, 2019. [Online]. Available: https://www.jetir.org/papers/JETIRDA06001.pdf

10. V. SRIKANTH, “Secured ranked keyword search over encrypted data on cloud,” IJIEMR Transactions, vol. 07, no. 02, pp. 111–119, Feb. 2018, [Online]. Available: https://www.ijiemr.org/public/uploads/paper/1121_approvedpaper.pdf

11. V. SRIKANTH, “A NOVEL METHOD FOR BUG DETECTION TECHNIQUES USING INSTANCE SELECTION AND FEATURE SELECTION,” IJIEMR Transactions, vol. 06, no. 12, pp. 337–344, Dec. 2017, [Online]. Available: https://www.ijiemr.org/public/uploads/paper/976_approvedpaper.pdf

12 . SRIKANTH MCA, MTECH, MBA, “ANALYZING THE TWEETS AND DETECT TRAFFIC FROM TWITTER ANALYSIS,” Feb. 2017. [Online]. Available: http://ijmtarc.in/Papers/Current%20Papers/IJMTARC-170309.pdf

14 Srikanth, V. 2018. “Secret Sharing Algorithm Implementation on Single to Multi Cloud.” International Journal of Research 5 (01): 1036–41. https://journals.pen2print.org/index.php/ijr/article/view/11641/11021.

5. K. Meenendranath Reddy, et al. Design and Implementation of Robotic Arm for Pick and Place by Using Bluetooth Technology. No. 34, 16 June 2023, pp. 16–21, https://doi.org/10.55529/jeet.34.16.21. Accessed 20 Aug. 2023.

16. Babu, Dr P. Sankar, et al. “Intelligents Traffic Light Controller for Ambulance.” Journal of Image Processing and Intelligent Remote Sensing(JIPIRS) ISSN 2815-0953, vol. 3, no. 04, 19 July 2023, pp. 19–26, journal.hmjournals.com/index.php/JIPIRS/article/view/2425/2316, https://doi.org/10.55529/jipirs.34.19.26. Accessed 24 Aug. 2023.

17. S. Maddilety, et al. “Grid Synchronization Failure Detection on Sensing the Frequency and Voltage beyond the Ranges.” Journal of Energy Engineering and Thermodynamics, no. 35, 4 Aug. 2023, pp. 1–7, https://doi.org/10.55529/jeet.35.1.7. Accessed 2 May 2024.

18. K. Meenendranath Reddy, et al. Design and Implementation of Robotic Arm for Pick and Place by Using Bluetooth Technology. No. 34, 16 June 2023, pp. 16–21, https://doi.org/10.55529/jeet.34.16.21. Accessed 20 Aug. 2023