State-of-art approaches for review spammer detection: a survey

Rupesh Kumar Dewang[1] · Anil Kumar Singh[1]

Received: 22 September 2016 / Revised: 20 February 2017 / Accepted: 21 February 2017 /
Published online: 29 March 2017

Abstract E-commerce websites are now favourite for shopping comfortably at home without any burden of going to market. Their success depends upon the reviews written by the consumers who used particular products and subsequently shared their experiences with that product. The reviews also affects the buying decision of customer. Because of this reason the activity of fake reviews posting is increasing. The brand competitors of the product or the company itself may involve in posting fraud reviews to gain more profit. Such fraudulent reviews are spam review that badly affects the decision choice of the prospective consumer of the products. Many customers are misguided due to fake reviews. The person, who writes the fake reviews, is called the spammer. Identification of spammers is indirectly helpful in identifying whether the reviews are spam or not. The detection of review spammers is serious concern for the E-commerce business. To help researchers in this vibrant area, we present the state of art approaches for review spammer detection. This paper presents a comprehensive survey of the existing spammer detection approaches describing the features used for individual and group spammer detection, dataset summary with details of reviews, products and reviewers. The main aim of this paper is to provide a basic, comprehensive and comparative study of current research on detecting review spammer using machine learning techniques and give future directions. This paper also provides a concise summary of published research to help potential researchers in this area to innovate new techniques.

Keywords Review spam and spammer · Sentiment mining · Supervised and unsupervised technique · Review dataset etc.

✉ Rupesh Kumar Dewang
rupeshdewang@mnnit.ac.in

Anil Kumar Singh
ak@mnnit.ac.in

[1] Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India 211004

1 Introduction

Buying products online has become a common phenomenon in todays growing market. Many customer are using e-commerce websites for the same. Before buying any product, the customer goes through reviews available for the product. Nowadays, the reviews of products have become an important source of information for the buyers in their decision for purchase of goods. Because of this tendency of customers, online reviews have become a target for spammers. The spammers are people who write fake reviews for the products. These fake reviews have been referred to as opinion spamming by Mukherjee et al. (2011). For example, if a user writes all negative reviews for products of a brand and write all positive reviews for competing brand, this reviewer is clearly a spammer (Mukherjee et al. 2011). The companies have enlisted spammers to write positive fake reviews in order to elevate their product and subsequently write negative reviews to downgrade competitor of brand or company. Products with positive reviews attract more customers. Due to this attraction, the spammers have tried to write fake reviews to deliberately mislead the potential customers. Through the online shopping customers interact with sellers via websites, so they can take decision on the basis of written features of products.

The spammers even works in group for posting reviews. Such spammers are called as group spammers (Mukherjee et al. 2012). These spammers refer to a group of reviewers who worked together to write malicious reviews in order to promote or demote a set of target products. Group spammers are more dangerous than individual spammers because of their size. They can take control over the sentiments of the product and can easily misguide the customers. It has been observed that there are too mamy spam reviews in leading review websites, such as PriceGabber.com, ShopZilla.com, or Amazon.com etc. (Lim et al. 2010). There is no constraint on writing reviews and posting them to social media. Everyone is free to write anything. Because of this, certain vendors or products provider are taking advantage of promoting their product or demoting competitors product unfairly. The first reported work on detection of fake review and reviewers is by Jindal and Liu (2008). They called this as a *opinion spam detection*.

The online review system has become corrupted because of spammers and there is an urgent need for automatically detecting spammers. Unlike web spam or email spam, review spam is much assiduous to detect. The key logic is that spammers can act as genuine reviewers. They simply impersonate themselves and it is hard for a human as well as machine to detect them. The combination of individual and group spammers have dramatically acted upon the accuracy of online reviews and this consequently raises concerns regarding the trustworthiness of the review system. In past few years, researchers have developed some spam and spammer detection techniques to preserve the accuracy of online reviews.

1.1 Earlier reviews

In year 2015, Heydari et al. (2015) did the first survey on review spam detection. They discussed the various existing papers in review spam detection. The majority terms used in paper for extract on review spam detection techniques. The data categorization is based on content of review, meta-data of review and information of product. Authors have discussed the results on the basis of problems used by researchers for review spam and spammer detection and focused on review duplication detection, content based methods, genre identification etc.

Another survey paper is recently published by Crawford et al. (2015) in 2015–16. They mainly focused on review spam detection using machine learning technique. The main purpose of paper is to explore the effects of Big Data analytics in review spam detection. The authors have used features description terminology for the discussion of papers. They have done very good comparative analysis of existing review spam detection techniques.

Alongside with these reviewed papers, larger number of work has accounted from year 2012 by the researchers for review spam and spammer detection . The various factors are included by the e-commerce company to stop the spam review posting. For example, single E-mail id should be used for account opening which is linked with mobile phone number. The verification code for opening account must be send on registered mobile number. Along with this, there are less explored areas by researchers such as, tagging of review spam and review spammer on e-commerce websites. There is very limited work done by researchers in spammer detection. There is only one survey paper published on spammer by Heydari et al. (2015), in which in depth detail about the review spammer is not discussed.

This review paper varies from existing writing overview in different way:

1. We classified and presented the review spammer detection features (group spammer features and individual spammer features) used in existing papers with detailed explanation in Tables 1 and 2. The analysis of the impact of features is done to identify the role of features in detection of spammers.
2. This paper presents the taxonomy of review spammer detection approaches classified in supervised, unsupervised and hybrid categories which is further classified into sub-parts with detailed discussion of proposed methods used in existing papers.
3. We have done comparative analysis of surveyed articles on the basis of “Performance Metric, Kappa Static Results, Feature Used for Model Building, Results/Score In Percentage, Performance Validation Method Used which is presented in Table 4. The pros & cons of each reviewed paper is presented. The comparative study of different technique presented in order to assess and find the most efficient technique for spammer detection.
4. Table 3 shows detailed of Review Dataset Summary Used By Referenced Papers.
5. It provides insights and future direction in detection of spammers for researchers.

In this way, work tending to these issues are considered for this review.

Rest of this paper is structured as follows: Section 2, introduces the history of spamming; Section 3 discusses the organization and methodology of survey; Section 4 presents the various features used in spammer detection; Section 5 review spammers detection techniques is presented. The Section 6 presents the discussions. Section 7 presents the future works and in last Section conclusion are presented.

2 History of spamming

Spam is basically defined as, distorted message sent using electronic media to numerous recipients. The messages may also contain some secret script to perform a cyber attack by cyber criminals. The work in spam detection started with e-mails and then similarly in numerous other media for example, chat spam, newsgroup spam, web spam, blog spam, SMS spam, Internet forum spam, social spam, file sharing spam and review spam.[1] The first email spam sent by DEC[2] in 1978. The DEC company was impacted negatively and

[1] https://en.wikipedia.org/wiki/Spamming

[2] http://www.templetons.com/brad/spamreact.html

Table 1 Group spammer features

S.no.	Feature	Authors	Year
1	Group Time Window (GTW)	[Mukherjee et al. Xu et al.]	(2011, 2012) (2013)
2	Group Deviation (GD)	[Mukherjee et al. Choo et al. Xu et al.]	(2011) (2015) (2013)
3	Group Content Similarity (GCS)	[Mukherjee et al. Choo et al. Xu et al.]	(2011) (2015) (2013)
4	Group Member Content Similarity (GMCS)	[Mukherjee et al.]	(2011, 2012)
5	Group Early Time Frame (GETF)	[Mukherjee et al. Xu et al.]	(2011, 2012) (2013)
6	Group Size Ratio (GSR)	[Mukherjee et al. Xu et al.]	(2011, 2012) (2013)
7	Group Size (GS)	[Mukherjee et al. Xu et al.]	(2011, 2012) (2013)
8	Group Support Count (GSUP)	[Mukherjee et al. Xu et al.]	(2011, 2012) (2013)
9	Maximum One Day Review Ratio(MOR)	[Choo et al.]	(2015)
10	Review Burstiness (BST)	[Choo et al.]	(2015)
11	First Review Ratio	[Choo et al.]	(2015)
12	Deviated Rating Ratio	[Choo et al.]	(2015)
13	Rating Abuse item Ratio(RA)	[Choo et al.]	(2015)

personal users were mostly effected. The users precious time was wasted to recognize which mails are non-spam and which are spam (Hinde 2002). Later on, the research direction was changed to filter out these types of spam mails (Sahami et al. 1998; Li et al. 2006). Li et al. (2006) defined the new filtering method for email spam by W-voting. They used content obscuring terms in spam message.

Next category is web page spam, which is defined as unauthorized way to stimulate the rank and orientation of web page. Spammer increases the ranking of web pages to attract the web users to visit these pages (Gyongyi and Garcia-Molina 2005). Review spam is identical to web page spam. Web page spam is commonly explained in Baeza-Yates et al. (2005), Wu and Davison (2005), Wu et al. (2006), Luckner et al. (2014), Vorakulpipat et al. (2012). Newsgroup spam[3] is an openly available network on the web. Newsgroup spam also provided group talks and group email messaging. In newsgroup spam, user or spammer posts advertisement. The aim of spammers is to target the reader of newsgroups (Choudhury et al. 2005).

Blog spam is form of the search engine spam. Blogs spamming is done through posting random messages, guest-book, wikis and other publicly-accessible online discussion boards.

[3] https://en.wikipedia.org/wiki/Spamming

Table 2 Individual spammer features

S.no.	Feature	Author name	Year
1.	Individual Rating Deviation (IRD)	[Mukherjee et al.]	(2012)
2.	Individual Content Similarity (ICS)	[Mukherjee et al.]	(2012)
3.	Individual Early Time Frame (IETF)	[Mukherjee et al.]	(2012)
4.	Individual Member Coupling in a group (IMC)	[Mukherjee et al.]	(2012)
5.	Review Count (RC)	[Xu et al.]	(2013)
6.	Brand Deviation Score (BDS)	[Xu et al.]	(2013)
7.	Targeting Products (TP)	[Xu et al.]	(2013)
8.	Targeting Product Groups (TPG)	[Xu et al.]	(2013)
9.	General Deviation (GD)	[Peng, Xu et al., Lim et al., Jiang et al.]	(2014, 2013) (2010, 2013)
10.	Early Deviation (ED)	[Peng, Xu et al., Lim et al., Jiang et al.]	(2014, 2013) (2010, 2013)
11.	Content Similarity (CS)/Review Similarity	[Aye and Oo, Mukherjee et al., Kim et al., Fei et al.]	(2014, 2013) (2014, 2013)
12.	Maximum Number of Reviews (MNR)	[Mukherjee et al.]	(2013)
13.	Reviewing Burstiness (BST)	[Mukherjee et al., Fei et al.]	(2013, 2013)
14.	Ratio of First Reviews (RFR)/ First Product Review.	[Mukherjee et al., Lu et al., Liang et al.]	(2013, 2013) (2014)
15.	Standard Deviation (SDN)	[Liang et al.]	(2014)
16.	Ratio of Outlier Products (RO)	[Liang et al.]	(2014)
17.	Average Standard Deviation (ASDR) / Absolute Rating Difference	[Liang et al., Lu et al.]	(2014, 2013)
18.	Average Unhelpful Feedbacks (ANH)	[Liang et al.]	(2014)
19.	Average ratio Unhelpful Feedbacks (ARH)	[Liang et al.]	(2014)
20	Binary Features (BF)	[Liang et al.]	(2014)
21	Total Helpful Feedback Number and Rate	[Lu et al.]	(2013)
22	Minimum Time Interval	[Lu et al.]	(2013)
23	Reviewer Rating Difference	[Lu et al.]	(2013)
24	Review Number	[Lu et al.]	(2013)
25	Ratio of Amazon Verified Purchase (RAVP)	[Fei et al.]	(2013)
26	Rating Deviation (RD)/ Rating Spamming/ Rating Abuse/ Similarity	[Lim et al., Wang et al., Aye and Oo, Mukherjee et al., Lu et al., Fei et al., Jiang et al., Xu et al.]	(2010, 2012) (2014) (2013) (2013, 2013) (2013, 2013)
27	Burst Review Ratio (BRR)	[Fei et al.]	(2013)

Link spam is usually posted on web application, which is accepted and shown the hyper-link submitted by the users (Seneviratne et al. 2015; Berger et al. 2015). The SMS spam works on mobile phone; spammer sends many hyper-linked and fake messages on mobile phones.

Table 2 (continued)

S.No.	Feature	Author Name	Year
28	Review Text Spamming	[Lim et al. Mukherjee et al. Jiang et al.]	(2010) (2013) (2013)
29	Single product Group Multiple High Rating/Low Rating	[Lim et al., Wang and Liang, Mukherjee et al., Jiang et al.]	(2010) (2013) (2013) (2013)
30	Sentiment Score	[Peng]	(2014)
31	Review Word Length Score	[Aye and Oo]	(2014)
32	Early Time Frame	[Mukherjee et al.]	(2013)
33	Trustiness of Reviewers	[Wang et al.]	(2011, 2012)
34	Honesty of Reviewers	[Wang et al.]	(2011, 2012)
35	Reliability of Store	[Wang et al.]	(2011, 2012)
36	Duplicate/ Near Duplicate Reviews (DUP)	[Mukhrjee et al.]	(2013)
37	Extreme Rating (EXT) by the Reviewers	[Mukhrjee et al.]	(2013)
38	Positive/Negative word length score by the Reviewers	[Aye and Oo]	(2014)
39	Helpful feedback Rank Number	[Lu et al.]	(2013)
40	Length of Review	[Lu et al.]	(2013)
41	Sentiment of Review by the Reviewers	[Lu et al.]	(2013)
42	Average Rating	[Lu et al.]	(2013)

Other types of spam like social spam, file sharing spam and image spam are generated through phishing frauds (Carpinter and Hunt 2006). The main objective of phishing attack is to trick recipients for divulging sensitive information, such as bank account numbers, passwords (internet banking and mobile banking) and credit card details. In the year 2008, the existence of review spam defined by Jindal and Liu (2008). According to author, review spam is new type of spam which lies on E-commerce website in the form of reviews. They have defined three types of review spam:

- **Type 1 (untruthful opinions)**: when review is not given as a product alike, means review given for promoting and demoting the product reputation by positive and negative posting reviews.
- **Type 2 (reviews on brands only)** : When the review is written in consideration of only brands name, instance of product of that brands, spammers want to damage the reputation of brands.
- **Type 3 (non-reviews)**: when the review only contains advertisement and other irrelevant opinion of the products like question and answering and random text.

Based on above categorization, authors have detected, whether review is spam or not. They have also defined the existence of reviews spammer in social media. The detection of review spammer was started by Lim et al. (2010).

3 Organization and methodology of survey

The first paper was published for detection of reviewers spammer in 2010 by, Lim et al. (2010). Jindal and Liu (2008), Jindal and Liu (2007) published first paper for detection of reviews spam.

Mukherjee et al. (2012) has shown the concept of groups spammer first time in 2012.

This survey paper has presented three main dimensions. The first dimension is spammer detection features which are further sub classified into group spammer features and individual spammer features. Tables 1 and 2 presented various features found in existing literature on review spammer detection .

These features are very useful for further research in reviews spam and spammer detection. In second dimension, we have classified the machine learning techniques used for detection of review spammer shown in Fig. 4. The machine learning techniques are classified into supervised techniques, unsupervised technique and hybrid techniques. The supervised techniques are further divided into two categories: graph based and linear classifier, in which graph based is further divided into probabilistic graphical model and other type of models and in same way linear classifier is divided into logistic regression and support vector machine (SVM). The unsupervised techniques is further divided into Bayesian network method, and clustering method. Hybrid method is further divided into learning to rank methods. In, Sections 5.1–5.3 we have discussed different tasks and sub-tasks along with applied approaches and techniques. In Table 3, we have presented the analysis of datasets which is used in existing papers. Table 4 shows the compression among in all existing paper presented in this survey paper. Table 5 shows the pros and cons of each paper. In discussion section we have provided the details discussion of all presented table (Tables 1 to 5). The open problems and limitations in research area is discussed in Section 6.

The reviews spammer are generally of two types: group spammer and individual spammer as shown in Fig. 1. In general, a person who writes non-genuine reviews is an individual spammer. If a spammer works with other spammers in writing spam reviews, then the spammers are called group spammer.

Figure 2 shows the group spammer example (Mukherjee et al. 2012), which is holding very suspicious patterns like, (a) all three members have reviewed same three products and all gave five star ratings; (b) all posted reviews with small time frame (4 days); (c) all reviews are posted by reviewers very early (when product was lunched). All above patterns occurring together denotes the suspicious activities because,we have seen all three reviewers as individual, so all reviewers seems genuine. All reviews received highest helpfulness votes from amazon users.

Figure 3 represents the general approach reviewer spam detection process. The first step is to select a review dataset. The review dataset may contain many impurities, so in next step pre-processing of datasets is performed. In next step, new features extraction is performed. The new features are constructed in the feature construction phase. Once all the features are computed then most relevant features for the task are selected in the next step of feature selection. Based on the selected features a classification model is built in model building phase. This model is then used for classifying the reviewer as spammer or non-spammer.

Table 3 Review dataset summary used in referenced papers

S.no	Dataset Name	Source	Number of reviews	Number of reviewers	Number of produst/stores	Author Name	Year
1	M-products(2010)	Amazon.com	109,518	53,469	39,392	[Mukherjee et al.]	(2012)
2	Store reviews	Resellerratings.com	408,470	343,603	1456	[Wang et al.]	(2011, 2012)
3	Product reviews	Datatang.com	10020	291	9384	[Jiang et al.]	(2013)
4	Restaurents reviews	Dianping.com	493,982	206,586	278	[Huang et al.]	(2013)
5	Products reviews (DVD, Music)	Amazon.com	1703	65,098	53,353	[Wang and Liang]	(2013)
6	Software category	Amazon.com	210,761	50,704	112,953	[Fei et al.]	(2013)
7	Electronic product	Amazon.com	195,174	141,501	300,864	[Lu et al.]	(2013)
8	M-products(2012)	Amazon.com	985,765	50,704	112,055	[Mukherjee et al.]	(2013)
9	M-products(2012)	Amazon.com	1,205,125	645,072	136,785	[Xu et al.]	(2013)
10	Product review	Amazon.com	3,794,694	1,037,621	962,234	[Aye and Oo]	(2014)
11	Product reviews (Music,DVD,Book)	Amazon.com	226,764	144,401	36,254	[Liang et al.]	(2014)
12	Store reviews	Resellarratings.com	628,707	561,703	8737	[Peng]	(2014)
13(a)	Book	Amazon.com	620,131	70,784	116,044	[Choo et al.]	(2015)
13(b)	Movie	Amazon.com	646,675	273,088	48,212	[Choo et al.]	(2015)
13(c)	Electronics	Amazon.com	542,085	424,519	35,992	[Choo et al.]	(2015)
13(d)	Tools	Amazon.com	229,794	151,642	22,019	[Choo et al.]	(2015)
13(e)	Across	Amazon.com	203,8685	901812	222,267	[Choo et al.]	(2015)
14(a)	Place-Centric	TripAdvisor.com	229,794	151,642	22,019	[Fayazbakhsh and Sinha]	(2012)
14(b)	Review-Centric	TripAdvisor.com	27,952	1,000	20,622	[Fayazbakhsh and Sinha]	(2012)
15	Software Marketplace (SWM)	anonymous site	1,132,373	966,842	15,094	[Akoglu et al.]	(2013)
16	NBA dataset for XDATA 2014	Yahoo/espn sport website	352,936	423,82	34,81	[Kim et al.]	(2014)
17	M-products	Amazon.com	—	2873	—	[Lim et al.]	(2010)

Table 4 Comparative analysis of existing works on the basis of Performance Metric, Kappa Static Results, Feature Used for Model Building, Results/Score In Percentage and Performance Validation Method Used

Author Name	Year	Performance Metric	Kappa Static Results	Feature Used for Model Building	Results Score in Percentage	Performance Validation Method
[Lim et al.]	(2010)	Rank NDCG	Substantial Agreement	Rating, Review Text & Product Groups	94 %	Human Evaluation & Rank Based
[Huang et al.]	(2013)	Rank NDCG	Substantial and Perfect Agreement	Sentiment Lexicon & Review Posting Frequency	98 %	Human Evaluation & Rank Based
[Peng]	(2014)	Rank NDCG	Substantial Agreement	Sentiment Analysis, Rating & Sentiment based Deviation Analysis	94 %	Human Evaluation & Rank Based
[Xu et al.]	(2013)	F-1 Score, MCC(Matthews Correlation Coefficient)	—	Review (Linguistic Indicator), Reviewer Behavior, Rating & Product Group review Behavior	95 %	Clustering & Graph based
[Fei et al.]	(2013)	Accuracy	Substantial Agreement	Review Posting time Frame, Rating & Review Similarities	78 %	Human Evaluation & K-means
[Wang et.al]	(2011, 2012)	Precision and Consistency	Substantial Agreement	Review, Reviewer & Store Features	70 %	Human Evaluation & Information Retrieval
[Fayazbakhsh and Sinha]	(2012)	LOF Score	Substantial Agreement	Rating, Review Time posting & Helpfulness Feedback Time	89 %	Human Evaluation & Density based
[Lu et al.]	(2013)	Accuracy	—	Review, Reviewer & Product	88 % & 90 %	SVM, LR & CRF
[Choo et. al]	(2015)	ROC Curve	—	Sentiment Analysis, Review Text, Rating & Review Posting Frequency	90 %	User statics, Verified purchased & Spamming Analysis

Table 4 (continued)

Author Name	Year	Performance Metric	Kappa Static Results	Feature Used for Model Building	Results Score in Percentage	Performance Validation Method
[Akoglu et al.]	(2013)	Rank and Ratings	—	Product, Review, Rating & Review Posting Frequency	75 %	Competitor Comparison
[Wang and Liang]	(2013)	Reliability Matrix	–	Rating & Reviewer		Cluster NMF
[Mukherjee et al.]	(2013)	Accuracy	Substantial Agreement	Author Features, Review, Rating & Review Posting Frequency	76 %	Human Evaluation, SVM, Rank boost & Posterior Analysis
[Mukherjee et al.]	(2012)	NDCG and AUC	—	Review, Reviewer, Rating Review Posting Time Window	95 %	Ranking Experiment & Classification Experiment
[Mukherjee et al.]	(2011)	SVM Rank	Perfect Agreement	Review, Reviewer, Rating Review Posting Time Window	92 %	Human Evaluation, Ranking &SVM
[Jiang et al.]	(2013)	NDCG ranking	Substantial Agreement	Reviewer, Review, Product, Rating & Review Posting Frequency	94 %	NDCG & Human Evaluation
[Aye and Oo]	(2014)	Accuracy	Substantial Agreement	Review, Reviewer & Rating	90 %	Human Evaluation & Ranking
[Kim et al.]	(2014)	Density and Ranking		Ranting & review Similarities		SVD
[Liang et al.]	(2014)	Precision	Substantial Agreement	Rating, Reviewer & product	64%	IR based
[Sandulescu and Ester]	(2015)	F1-score and Precision	–	Review Similarities, Topic in Review & Semantic Similarities	65 %	Bag-of-word Model

Table 5 Pros and cons of existing papers

Author Name	Year	Pros	Cons
[Mukherjee et al.]	(2011)	1. Used co-clustering model. 2. Non-negative tri-factorization algorithms used.	1. Evaluation of reviewer reliability required. 1. Complex algorithm for validation on real time larger dataset.
[Mukherjee et al.]	(2012)	1. Define new eight group spammer indicators. 2. Used Spamicity to cluster the spammer or non-spammer group.	1. Manually label dataset is high cost. 2. Not possible to applied real time tagging.
[Xu et al.]	(2013)	1. Behavioural features used. 2. Find near duplicate review. 3. Graph bound the same reviewer post review in within same time span.	1. Point-wise features used. 2. Required real fake review model training. 3. Assumption of one reviewer posting many review is not realistic.
[Choo et al.]	(2015)	1. Model builds on user interaction. 2. Sentiment analysis performed on the basis of votes and helpful feedback. 3. Strong positive communities build for detection.	1. Review content has not used for detection. 2. Individual non-group spammer not find.
[Lim et al.]	(2010)	1. Model has used centric and user behavior driven features. 2. Target based method used.	1. Depends on temporal and behavioral features. 2. Proposed algorithm heuristics in nature. 3. Reviews inspected Manually is time consuming and costly endeavor.
[Wang et al.]	(2011, 2012)	1. Randomly initiates scores and work initerative manner. 2. Model includes more hidden spamming clues.	1. Proposed model is not generalize, and Store quality changed over time. 2. less domestic study.

Table 5 (continued)

Author Name	Year	Pros	Cons
[Fayazbakhsh and Sinha]	(2012)	1. Algorithm works in non-iterative nature. 2. Suspicious score has calculated for review, reviewer & products.	1. Human evaluation method has not defined properly. 2. Proposed model is not generalized to all dataset.
[Jiang et al.]	(2013)	1. Voting based model proposed. 2. Spammer behavior scoring used.	1. Manually labeling takes times & high cost. 2. Only activity model is new other things are already used in previous research.
[Huang et al.]	(2013)	1. Linguistic & behavior of review is used. 2. Sentiment based model proposed.	1. Proposed algorithm is heuristics in nature. 2. Only small sample taken for labeling. 3. Used dataset is language depended.
[Fei et al.]	(2013)	1. Used new Amazon Verified purchase Features. 2. Hidden node concept used to repress the reviewer real identity.	1. Used temporal and static features. 2. Model worked on limited number of datasets.
[Lu et al.]	(2013)	1. Untied detection framework proposed. 2. Annotators of review has performed.	1. Manually labeling of dataset is time consuming. 2. There is not any method to verify the accuracy of human evaluated spammer or non-spammer.
[Mukherjee et al.]	(2013)	1. Credibility calculation performed. 2. Model works in smaller time window. 3. Used reviewer behaviors.	1. Manually label dataset is high cost. 2. Heuristics in nature. 3. Spammer easily changed content based detection model.
. [Akoglu et al.]	(2013)	1. Establish correlations between reviewers and stores. 2. Proposed model signed Nature.	1. Reviewer interaction Patterns are quite different from regular users. 2. Proposed model has not used dynamic labeled anomalies.

Table 5 (continued)

Author Name	Year	Pros	Cons
[Aye and Oo]	(2014)	1. Reviewer behavior based scoring used. 2. Text retrieval conferences (TERC) is used to measure the performance of proposed model.	1. Not consider the deep understanding of review text.
[Kim et al.]	(2014)	1. Content similarity calculated by the structural rank. 2. Bipartite graph algorithm is faster. 3. Proposed Trust-Rank the semi-automatic model.	1. Need additional attention measuring in multi-documnet similarity
[Liang et al.]	(2014)	1. Graph characteristics are integrated with reviewers features and spamming behaviour.	1. Not focused on the text or content of reviews.
[Peng]	(2014)	1. Relationship based model proposed. 2. Single & multi-mode spammer detection is performed.	1. Manual labeling of dataset has performed on the basis of30 ways report, which is very hard to apply perfectly by all judges. 2. Sentiment & rating co-relation is harder to achieved in real time dataset.

Fig. 1 Review spammer types

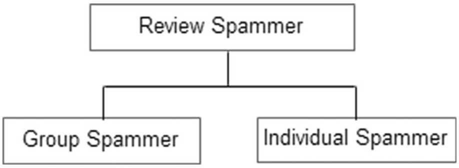
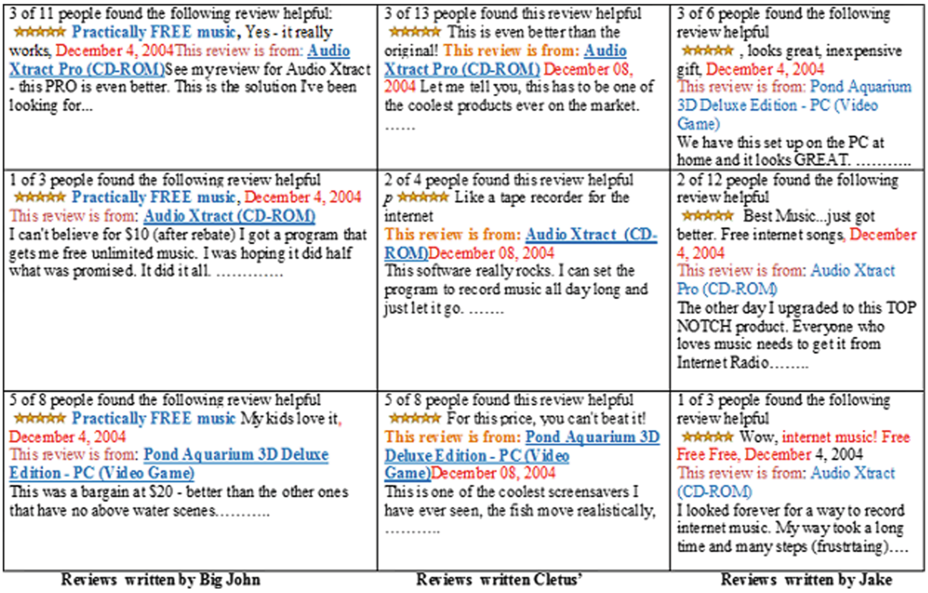


Figure 4 elaborates the pre-processing steps for reviews and reviewers. The Pre-processing of reviews involves mainly stop-words removal, POS tagging (part of speech), tokenization, stemming and lemmatization etc. and for reviewer, pre-processing contain the removal of anonymous user, removal of duplicate products and removal of inactive users or unpopular products. The pre-processing of dataset is first essential task before features extraction.

In Table 3, we have complied the details of review datasets used in existing work. Many papers have used Amazon.com review dataset. Some papers, Lim et al. (2010), Wang et al. (2011), Liang et al. (2014) etc. took review datset from Resellerratings.com. Wang et al. (2012) used dataset from Dianping.com and in paper Aye and Oo (2014) NBA dataset is used.

4 Feature used in spammer detection

There are many features defined and constructed by various authors pertaining to the detection of review spammer. Review datasets are largely unstructured texts. Representation of review datasets is very important. One of the famous representation is bag-of-words (BoW), also called vector space model. Spammer features are categorized into two types:



3 of 11 people found the following review helpful: ★★★★★ Practically FREE music, Yes - it really works, December 4, 2004This review is from: Audio Xtract Pro (CD-ROM)See myreview for Audio Xtract - this PRO is even better. This is the solution I've been looking for...	3 of 13 people found this review helpful ★★★★★ This is even better than the original! This review is from: Audio Xtract Pro (CD-ROM) December 08, 2004 Let me tell you, this has to be one of the coolest products ever on the market.	3 of 6 people found the following review helpful ★★★★★ , looks great, inexpensive gift, December 4, 2004 This review is from: Pond Aquarium 3D Deluxe Edition - PC (Video Game) We have this set up on the PC at home and it looks GREAT. ....
1 of 3 people found the following review helpful ★★★★★ Practically FREE music, December 4, 2004 This review is from: Audio Xtract (CD-ROM) I can't believe for $10 (after rebate) I got a program that gets me free unlimited music. I was hoping it did half what was promised. It did it all.	2 of 4 people found this review helpful *p* ★★★★★ Like a tape recorder for the internet This review is from: Audio Xtract (CD-ROM)December 08, 2004 This software really rocks. I can set the program to record music all day long and just let it go.	2 of 12 people found the following review helpful ★★★★★ Best Music...just got better. Free internet songs, December 4, 2004 This review is from: Audio Xtract Pro (CD-ROM) The other day I upgraded to this TOP NOTCH product. Everyone who loves music needs to get it from Internet Radio........
5 of 8 people found the following review helpful ★★★★★ Practically FREE music My kids love it, December 4, 2004 This review is from: Pond Aquarium 3D Deluxe Edition - PC (Video Game) This was a bargain at $20 - better than the other ones that have no above water scenes...........	5 of 8 people found this review helpful ★★★★★ For this price, you can't beat it! This review is from: Pond Aquarium 3D Deluxe Edition - PC (Video Game)December 08, 2004 This is one of the coolest screensavers I have ever seen, the fish move realistically, ...	1 of 3 people found the following review helpful ★★★★★ Wow, internet music! Free Free Free, December 4, 2004 This review is from: Audio Xtract (CD-ROM) I looked forever for a way to record internet music. My way took a long time and many steps (frustrtaing)....
Reviews written by Big John	**Reviews written Cletus'**	**Reviews written by Jake**

Fig. 2 Group spammer example (Mukherjee et al. 2012)

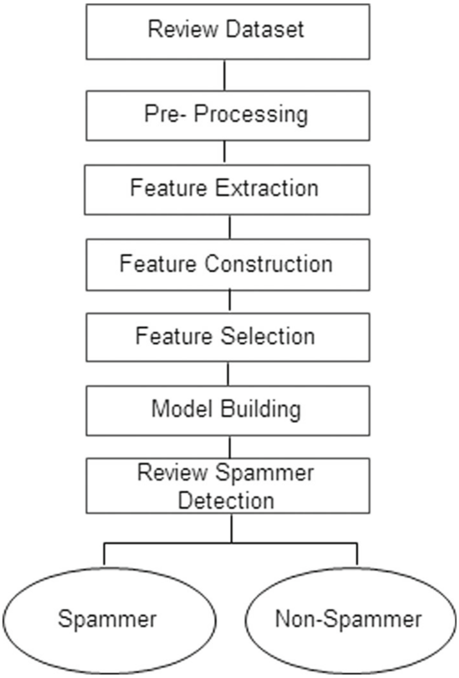


Fig. 3 Review spammer detection process

4.1 Group spammer features

Feature used for describing or defining to detect group spammers is called group spammer features. Table 1 shows the features name and definition of features used in previous researches (some formulas are also defined by authors for particular features, due to space complication they are not shown in this survey) and last column show references. There are total thirteen numbers of features used. Mukherjee et al. (2011, 2012, 2013), Xu et al. (2013), Choo et al. (2015) have defined some very interesting features for detection of

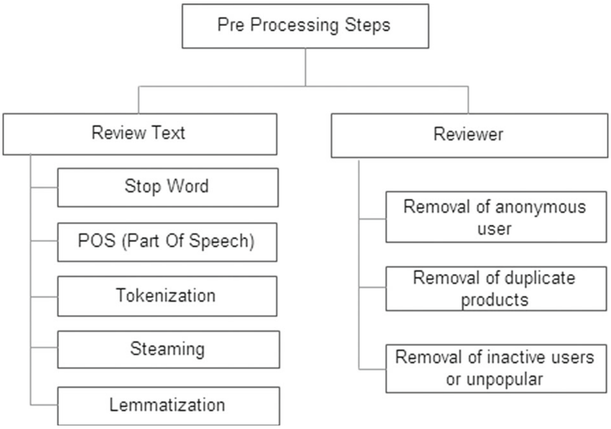


Fig. 4 Preprocessing steps

group spammers, some of which are: “group early time frame” which shows the behavior of spammers for promoting just launched products. When any products are newly launched in market, spammer gain more advantage by posting biased reviews within short time interval. If spammers copy the other users review word by word and post then, this feature is known as “group content similarity”. This feature is very useful for similarity measure of group content.

To the best of our knowledge, up till now there are only five papers published for detection of group spammers. In which first one was published by Mukherjee et al. (2011, 2012). The latest paper regarding group spammer features is published in 2015 (Choo et al. 2015).

4.2 Individual spammer features

The features that describes the characteristics of an individual spammer is called Individual spammer features. Individual spammers work alone for writing fake reviews. Sometimes, they just copy exact text of other posted review for review writing. To the best of our knowledge up till now, there are fifteen papers published for detection of individual spammer in which first one is by Lim et al. (2010). In Table 2 these feature are described in terms of feature name, its definition and reference number of papers. Some important features are: Individual Rating Deviation (IRD) used to identify biased higher rating given by spammers. This higher rating is deviated from general rating given by the other users (Mukherjee et al. 2012). Targeting product groups are used to measure the ratings of set of products (which share common attributes). The meaning of some features (General Deviation, Early Derivation, Content/Review Similarity etc.) and their formula are same in many published papers (Peng 2014; Xu et al. 2013; Lim et al. 2010; Mukherjee et al. 2013; Lu et al. 2013; Fei et al. 2013).

5 Review spammers detection techniques

The detection techniques of review spammers could be divided into three parts based on their usage of supervised, unsupervised techniques and other techniques. A survey of techniques used in various published papers is presented in next subsection. The presented division is based on the types of techniques used by researchers. In next subsection, we have covered the details explanation of existing articles on the basis of Fig. 5.

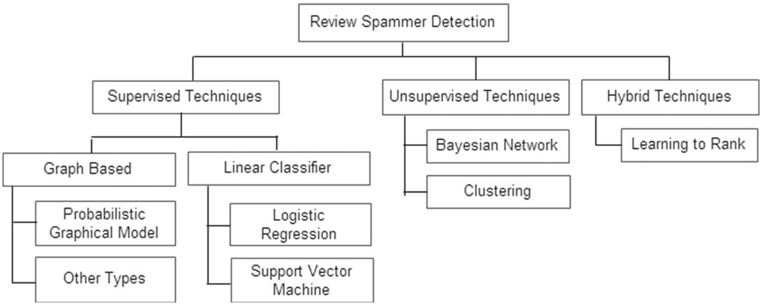


Fig. 5 Taxonomy of state-of-art approaches for review spammer detection

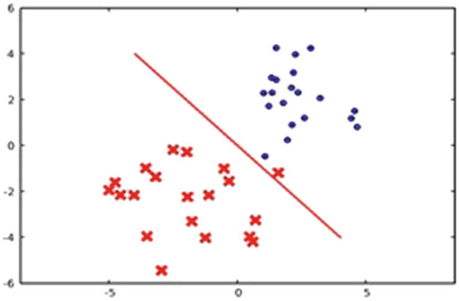


Fig. 6 Classifying point by straight line (Yuan et al. 2012)

5.1 Supervised technique

Supervised learning is a method of inferring a function from a labeled data. Training dataset is used for learning of model and test datasets are used to classify unseen records. There are various algorithms for the supervised classification. In next subsection, we have described various classification methods used by different researchers.

5.1.1 Linear classifier

It is a type of supervised learning. Linear classifiers uses a of linear combinations of properties and features for building classification decisions (Yuan et al. 2012). Figure 6, represents two types of objects ‘x’ and ‘O’. In this figure, the objects x’s and O’s are partitioned with a straight line. So the linear classification for straight line is calculated as;

Given a single point X = (x,y), we can express the classification of the point as: Sign (ax+by+c), where a, b and c are constants that define the line. This function will return a+1 if ax+by+c is positive. Linear classifier is further classified into logistic regression based method and Support Vector Machine (SVM) which are explained next.

Logistic regression based methods Logistic regression estimates the probability of an event occurring. Logistic regression is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome.[4]

The logistic regression based method was first used in Lim et al. (2010) for detection of review spammer. They proposed a scoring technique method to find the spamicity degree of each reviewer. The authors show that the detection methods are based on several predefined abnormalities indicators, such as general rating deviation, early deviation - i.e. how soon a product appears on the website, does a suspicious user post a review about it or very high/low ratings clusters. Proposed methods are user centric and user behavior driven. They calculated an overall spam score for a reviewer. For this purpose, further a linear weighted combination of factors for measuring spamming behavior is used. The preprocessing for reviewers are performed to eliminate noisy data.

A novel spamming behavior model based on pattern of review content and rating of reviews for particular products is proposed. The models *targeting products (TP) and targeting product groups (TPG)* are called the target based spamming and *group deviation(GD) and early deviation (ED)* are called the deviation based spamming.

[4] https://www.medcalc.org/manual/logistic_regression.php

- **Targeting Products (TP)**: Review spam is mainly written for targeting products and it is very easy to identify. For identifying targeting products they have covered three parameters: (a) Ratings spamming, when reviewer gives the highest rating to same type of pair of products. (b) Review text spamming: when two or more reviews are similar and spammer do some minor changes in old review and divide big review into small. (c) Combined spam score: authors have combined above two methods, to check whether the ratings lies within some decided threshold or not.
- **Targeting Product Groups (TPG)**: In this model the spammer changes the ratings of same brand products by sharing some common characteristics in very short interval of time. These are, (a) single products group multiple high ratings: When one single product group is targeted by the spammer to give multiple high ratings to products. (b) Single products group multiple low ratings: In this they decided just opposite behavior by assigning several low ratings. (c) Combined spam score: Authors have combined above two methods, to check whether the ratings lies within some decided threshold or not.
- **General Deviation (GD)**: When the spammer gives rating on the basis of promoting and demoting a particular product, then it is called the general deviation.
- **Early Deviation(ED)**: When the first time product is launched and the spammer use this opportunity by giving false rating of products. Such type of spamming behavior, manipulates the mindset of other reviewers.

The features weight were linearly combined towards a spamicity formula and computed empirically in order to maximize the value of the normalized discounted cumulative gain measure. Next, they executed the trained model to score reviewers. The measure shows how well particular rankings improved on overall goal. They also used human judgment for evaluating the method. The proposed approach is heuristic in nature. The manually labeling of deceptive reviews has proven to lead low accuracy while testing on real-life fake review data.

Very similar approach was proposed by Huang et al. (2013). They proposed a model for detecting professional spam reviewers. The proposed model analyze the writing and behavior styles of the spammers. They used linguistic feature of review and behavior of the review spammer, review posting frequency and reviews sentiment strength. The method for counting of reviews posting frequency on particular time has been introduced and named as Frequency Spammer Detection Model (FD). The reviews sentiment strength has been evaluated by counting the sentimental words and named as Sentimental Spammer Detection Model (SD). These approaches were further combined to form a linear combination model. The accuracy of the model is improved by large margin compared with method (Lim et al. 2010) .

Another approach proposed by Peng (2014), uses sentiment analysis techniques to calculate the sentiment score of reviews. Author identified the deviation in rating and sentiment by mapping the heterogeneous graph of reviews and stores. They proposed relationship-based method for identifying the spammers. The author has given an algorithm for detection of single-mode spammers and multi-mode spammers. For calculating the sentiment, they had generated sentiment lexicon using SentiwordNet (Esuli and Sebastiani 2006) and MPQA (Wilson et al. 2005). Experimental results shows that proposed method identified both single-mode and multi-mode spammers coherently with good precision.

Support Vector Machines classifiers (SVM) SVMs are generally used to determine linear separators in search space. SVM helps to isolate the non-identical classes. The nature of text is sparse and for that reason text data can by suitably used for SVM classification.

The text data generally contain small number of features that are immaterial, but they are harmonized with one another and are well ordered into linearly separable categories. SVMs are used in many applications, amongst these applications, classifying reviews according to their quality is one of them.

In review spammer detection, the SVM technique used for model building and bootstrapping purpose. Xu et al. (2013) has used SVM to compared with KNN (k-Nearest Neighbors) based and graph based proposed algorithms. In same way the SVM method is mainly used to compare with proposed algorithms by various authors: Mukherjee et al. (2011, 2012), Fei et al. (2013), and Lu et al. (2013).

5.1.2 Graph-based

Graph-based approach is the concept of representing problems in terms of graphs. Several graphical properties are henceforth implemented to detect the concerned problem. Graph-based approach has also been used in detecting spammers in Lim et al. (2010), Wang et al. (2011), Fei et al. (2013). Many authors have used graph for representing the relationship between the reviews, reviewer and products. Wang et al. (2011) has used graph to show the relations between the reviewers, reviews and store of products. The same representation has used by the Fayazbakhsh and Sinha (2012), Wang et al. (2012). Liang et al. (2014) used two graph models. One is same as explained above and second represents the new concept of supportive or conflict relationship in between the reviewers. If the reviewers give same ratings to products, this shows the supportive relation otherwise shows the conflict relation. Lu et al. (2013) has used the factor graph model in which the set of reviewers and reviews are represented by the undirected links. Each links between the reviewer and reviews shows the reviewer writes the reviews for the products.

Probabilistic graphical model Probabilistic graphical models modulated the graph theory and probability theory. The multivariate statistical models are useful for detection of review spammers. The multivariate statistical modeling is given by Zhou (2011). They provided a unified description of uncertainty using probability and complexity using the graphical model. The method is classified into two parts, directed graphical model and undirected graphical model. The most commonly used directed probabilistic graphical model is Bayesian Network. The undirected probabilistic graphical model used two methods first one is Gaussian graphical model and second is Markov random field. Directed graphical model is sub-classified into Conditional Probability Distribution.

Xu et al. (2013) proposed the graph model which is based on pair-wise markov network. They assigned two class labels one is positive and other is negative for colluder. They used KNN based technique. The KNN method is compared with graph based method in which graph based gave better results. Authors have used hybrid methods, for this reason time complexity is increased .

Fei et al. (2013) proposed the markov chain random field method, in which reviewer writes reviews in same burst. It is defined as co-occurrence. They used two type of node, hidden node represent the real still unknown identity, which shows the three stage of reviewers in terms of non-spammer, mixed and spammer and other node is observed node which is combined with hidden node shows the current state of data. They did not calculate the frequency of two reviewers appears together in burst time.

Other types The graph based method for the detection of individual spammer in store is first introduced by Wang et al. (2011, 2012). They proposed a concept of heterogeneous

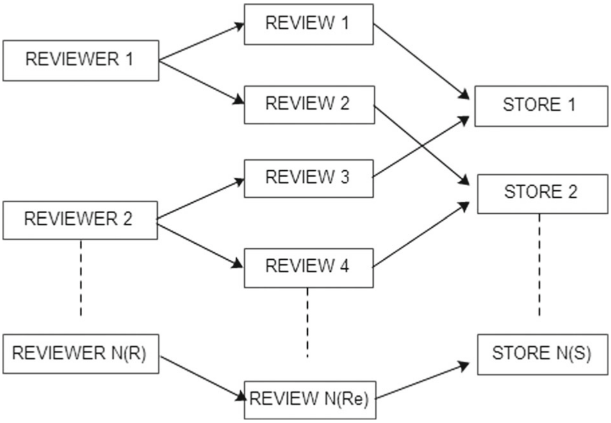


Fig. 7 Relationship between review, reviewer and store (Wang et al. 2011)

review graph for capturing the relationships between reviewers, reviews and stores, that the reviewers have reviewed in order to detect review spammers. There are three kinds of nodes in the review graph shown in Fig. 7. In this figure, if reviewer writes reviews for store, it is shown by links between reviewer, review number and store in which review is written. They identify trustworthiness, honesty and reliability among the reviewers, reviews and stores. The inter-relationship between reviews, reviewers and stores are also defined by the authors. The reinforcement based method is used for analysis of interrelationship.

The reviewer trustworthiness has defined on the basis of honest reviews. The honest reviews have posted by the reviewers on all products which represented the exact specification of products used. The stores are considered reliable when the trustworthy reviewers writes more positive reviews. The reviews are honest, if they are supported by many other honest reviews. The results shows that the methodology can recognize spamming activities with good value and human evaluator agreement.

In a similar approach used by Fayazbakhsh and Sinha (2012). Authors defined two stages of calculation for spammer detection; first one was suspicion score for each node in graph. Second, proposed iterative algorithm which contains two steps: forward updates and backward updates. The forward updates and backward updates have calculated the suspicion score based on graph connectivity. This is first method for detection of review spam using outliers detection technique.

Another graph based method has proposed by Lu et al. (2013). They proposed a Review Factor Graph Model to consolidate all the features to describe each review and reviewer and to create belief propagation between reviews and reviewers. In this paper, it expected to detect both the fake reviews and review spammers at the same time. Labeling of dataset has done by college students; for this they first gave training to students. They used four features; review related, reviewer related, feature between reviewer and review and review group feature. These all features are given in Table 1. An efficient max-sum algorithm is further designed to utilize belief propagation to perform model learning and inference. By leveraging the belief propagation between reviewers and reviews, the belief of reviewer and reviews can propagate and influence each other. Finally, they analyzed the contribution of each factor to detect fake reviews and review spammer based on the learned weights. The experimental results shows mutual enhancement effect, their approach achieved significant improvement over all other baseline methods such as SVM, Logistic Regression(LR) and Conditional Random field(CRF).

Recently researchers used the graph based method for group spammer detection in 2015 by Choo et al. (2015). They have proposed unsupervised hybrid approach which is based on user interactions coupled with sentiment analysis. They defined new type of spammer which is called Promotional opinion spammer, it refers to the spammer who is mainly using malicious artificial boosting for generating false opinions. They identified promotional opinion spammers through making user relationship. Spammers need high ranking which is collected mostly by positive response rather than negative ones. They used sentiment analysis of user interaction for finding out community structures to distinguish spam communities from non-spam. Using sentiment, they have build the positive relationship graph which is used to build the groups spammer identification. They shown that 80 percent of candidates are spammers. Author said that, this is the first method which identifies reviews spammer groups through analyzing interaction with users other than review text. The main advantage of this approach is, it detects review spammers purely on features other then review content-based method.

5.2 Unsupervised techniques

In unsupervised learning the data does not contain any target value. To find any native structures, one need to use data exploration techniques. Unsupervised learning does not use any class attribute, where as supervised learning require class attribute. It also requires the post-processing techniques to validate the generated results. The number of classes to be learned may also be not known in advance. We have shown bayesian network and clustering methods under unsupervised categorization below.

5.2.1 Bayesian Network (BN)

The Bayesian Network is unsupervised model in which, all the features are fully dependent. The Bayesian Network model is a directed acyclic graph. In this graph, random variables represent nodes and conditional dependencies represent edges. In the BN model, a complete joint probability distribution is specified over all the variables because it is considered a complete model for the variables and their relationships. Some recent works that used BN method for review spammers detection includes:

Akoglu et al. (2013) has proposed framework FRAUDEAGLE . It forms a network between the reviewers and products. The user scoring method is used for grouping different objects (products, reviews and reviewers) based on the classes. Each object is categorized into two classes; products are classified into bad or good quality, users are classified into honest or fraud and reviews are classified into real or fake. All above divisions is based on the inter-relationship between the reviews sentiment (positive or negative), writing style of user (positive or negative) for good or bad products and user behaviour (fake or real) for reviews. The authors has used ratings for analysis of sentiments of reviews. If a review rating is above the threshold value, it is considered as a positive, otherwise it is negative. The “sign symbols” are used for sentiment finding which are based on the thumps-up (positive) and thumps-down (negative). The fraud detection is based on the visualization and sense making methods. Authors have formulated the classification task of spam detection into network classification tasks. The representation of bipartite network is represented into user node with connection to products node. The authors claim that proposed method is scalable to large datasets and it is very general method, which can be applied all types of review networks. However the authors have missed the focus, on the problem setting, authors has not used review text and other information, such as time-stamp, product brand, etc.. The proposed assumption are not applicable in all conditions like, authors claim honest users always gives negative reviews to bad products or fraudsters always gives the positive reviews to good products vice-versa, it is not true on the real scenario when any competitor company hire spammers to writes negative reviews in very larger quantity to demolish the real image of products. As we know that, when users want to buy any products he or she looks bad or good things first and compares them with other company products reviews.

5.2.2 Clustering

Clustering is the task of grouping a set of objects in such a way that objects in the same group (or cluster) are more similar than objects in other group. It finds natural grouping of instances, when the given data is unlabeled. There are various clustering methods present like, k-means, hierarchical clustering and fuzzy clustering etc. In this survey, we have discussed previous papers based on clustering techniques.

Wang and Liang (2013) has proposed a data co-clustering methods for discovering ratings patterns. The computational study is performed on rating pattern of reviewers. The reliability of reviewers is calculated in the form of matrix which is called reliability matrix or heterogeneous relational data matrix (HRD). The reliability matrix is used in two data types, reviewer and product. The reliability matrix is formulated based on rating consistency. The rating consistency is defined to evaluate the trustworthiness of reviewer of a product. The nonnegative tri-factorization is used for reliability matrix and two types of clusters were extracted concurrently. A diverse relational data matrix (HRD) has constructed between a reviewer set and a product set. Here the central data type is the reviewer and the modality feature is the product. The proposed method bisects the reviewers but also made clusters of the products in mean time. Some challenges are presented like, the evaluation of reviewer reliability and the reliability matrix needed. The comprehensive description include more pertinent properties from textual comments, posting time, posting frequency, behavioral properties of neighbors and reviewed items and their features.

Mukherjee et al. (2013) has used fully Bayesian approach and translate the opinion spam detection as a clustering problem. Bayesian method has used to model the spamicity of reviewer on the basis of behavioral and latent features. The proposed model is Author Spamicity Model (ASM). They have not used any label data to evaluate the results. They used top and bottom ranked authors produced by the model as two separate classes to build the classifier. Authors also enable detection and posterior density analysis in a single framework. The model evaluation show that helpfulness votes are not only performed poorly but are also abused. The groups of spammers worked together to promote certain products may gives many votes to the reviews of each other. The main contributions are in reviewer spam detection and posterior density analysis in a single framework. They have used human expert evaluation for checking the correctness of proposed model. However in my opinion, authors have not used any review text features.

Xu et al. (2013) detected the collusive spammers in Chinese review websites. Frequent itemset mining method is used for forming the colluder (groups). Exact method has been used by Mukherjee et al. (2012). The annotations of reviewers in colluder or non-colluder are based on the Amazon.cn strategy to clearing the displayed reviews periodically. They select first most similar reviewers for voting. Authors have used KNN based clustering method. They proposed two methods for detecting collusive spammers: KNN based method and a general graph-based classification method. The KNN- based method used pair-wise similarity of reviewers. The pair-wise similarity of two reviewers is based on three major factors: common member ratio, common products ratio and common brand rating deviation. The distribution of pair-wise similarity scores over three types of reviewers pairs: negative-negative pairs, negative-positive pairs and positive-positive pairs.

The graph model is based on pairwise markov network. The reviewer has served at least one product in a fix time interval, so correlation is formed among the reviewers. Authors used approximate inference algorithm which is based on the iterative classification. This algorithm is designed for reviewers. The class label of reviewers is collectively determined by the intrinsic attributes and neighborhood reviewers. They used two methods, invariant conceal and behavioral histories of reviewers in relational data. It also performed empirical analysis on the three spam indicators; individual behavioral inductors, linguistic inductors and collusive behavioral inductors. The used collusive behavioral indicators and individual behavioral inductors has adopted from Mukherjee et al. (2012). Cumulative histograms has used to measures distributions of three inductors' scores in between the colluder or non-colluder. The anomalies have spotted not only on the basis of language colluders but also on the basis of their behaviors. The results shown that both the proposed methods are out-performed baseline classifiers on all indicator sets. However the proposed methods do not generalize to all data types, for example, in real time scenario reviewers only write single review for products, but in this paper author has ignored or eliminate single reviews in preprocessing steps. Only two indicators namely, brands name mentioned frequency (BNF) and squared average length (SAL) are defined.

5.3 Hybrid methods

In hybrid methods, many authors have combined more than two method for reviews spam detection. Mukherjee et al. (2011, 2012) has used rank learning with SVM and GS rank etc. Hybird method is further classified into Rank learning approaches which is shown below.

5.3.1 Rank learning approaches

Rank learning is a broad area in Information Retrieval (IR). It is machine learning application. Mainly supervised, semi-supervised and reinforcement learning are applications of machine learning. The model building steps are used in ranking algorithms for review spammer detection. It also helps to rank spammer and non spammer in dataset.

Mukherjee et al. (2012) used the hybrid method GSRank (Group Spam Rank) and several behavioral models based on individual reviewer, groups and products. Authors have proposed a frequent itemset mining technique to extract candidate groups of users. They have proved that label dataset of fake reviews which produced through crowd-sourcing methods is not valid. The crowd-sourcing models are not generalized well on real-life test data. They have done manual labeling of dataset by eight judges (judges from well known e-commerce company) on the basis of opinion spam signals. Eight judges labeling results are formalized by spamicity scores and multi-rater kappa methods. Authors have used measuring nominal scale agreement among many raters (Fleiss 1971) on eight indicators to find out the groups spammer, which is shown in Table 1. They have derived the relation of spam and non-spam using cumulative percentage between groups and features value. They found that non-spam groups features values are very low. The groups spammer worked with many other reviewers and written too many reviews about different products. We have given dataset table (Table 3) which shows the groups spammer dataset used by Mukherjee et al. (2011, 2012). Authors have used three combinations of group spammer detection model: groups spam-products model, members spam –products model and group spam-member model. GS-Rank method is used to check which entity affects other entity. Double checking of each model entities is the main advantage. The rank experiment and comparison are performed by using SVM-Rank, SVMRank-H, GSFSum, SVR, RankBosst, RankBoost-H and HS. Generated results shown that GS-Rank confidence level is 95 % when compared to all the above ranks methods. However the authors have not distinguished a single person with many user ids. Authors have ignored single review which is written by the single reviewer. This single review may be fake review. Similar approached used by Duh et al. (2013). Authors have added other features which are based on psycholinguist text analysis, but after this they did not propose any work.

Another paper by Jiang et al. (2013) is based on detection of spammers using Activity Model (Voting base approach). They proposed a scoring method to improve the accuracy of the original detection model. The main objective was to built a detection model to find the review spammers by analyzing the user behavior. This papers is extension of Lim et al. (2010). Authors have added two kinds of behaviors: First is *The pattern which reviews has continuously published in a short time. Second is, the pattern which numbers of reviews are more than the number of purchase.* Prior to this, the other three existing behaviors are used: First one is *the pattern in which same user published several reviews and ratings on the same product. Second is, the pattern in which same user published reviews and ratings for product groups. Third is, the pattern in which rating had a deviation from other reviewers.* The detection models based on the product group is divided into two: the high rating model and the low rating model. At amazon website, the highest score is 5 and the low score is 1 or 2. NDCG (Normalized Discounted Cumulative Gain) method is used to evaluate the model. The score is used as a standard to measure the spam. The reviewers who are getting the highest score will be the spammers. This paper improves the prediction model using behavioral approach and the results confirmed this idea. However, authors have not given any example of voting model. Authors have selected very small data for voting which will not be applicable in real dataset.

Another behavioral based method used in Aye and Oo (2014). The Proposed technique is based on reviewers spamming behaviors. This technique calculates spam scores based on the spamming behaviors of the reviewer. The review similarity is used to determine spammers. If the reviewer writes similar reviews then all detailed review behaviors and the rating spam score of the reviewer are calculated. Authors have assumed more spamming behaviors, when the reviewers get more spamming scores. The review similarity is calculated by using single method to detect suspicious reviewer instead of deep analysis on opinion mining. Review spam scores are calculated based on similar reviews and rating spam scores are calculated for more accurate results. The reviewers are more likely to be a spammer, if a review gets high spam score. Some spamming behaviors are stated as when two reviews are equal and most likely to be spam review. Other review and rating behaviors such as similarity, deviation and good review or bad review and posting date are also considered to detect spammer. The spammer detection system consists of three features: 1. Review Based Spam Score, 2. Rating Based Spam Score and 3. Combined scores. Prior to these 3 steps, authors calculated the review similarity for all reviews of reviewer. The results show that the presented technique has comparatively effective spammer detection success than other technique based on helpfulness vote alone. However in my opinion, limitation observed in this paper is that, the technique focused only on scoring methods and is not considered deep understanding of the review text. So there is a need to add some classification method to achieve more effective spammer detection system.

Kim et al. (2014) has proposed an efficient spammer detection method using structural rank of author specific term-document matrices. The structural rank is the maximum rank of all possible matrices of the same non-zero pattern. It only considered the non-zero patterns and used bipartite graph traversal algorithms. In this paper, structural rank has used for computing content similarity of a set of documents. The non-zero pattern is enough to measure the content similarity of a set of documents. The bipartite graph traversal algorithm is also being much faster than other pair-wise based similarity matrices. A Spammer Scores feature is proposed which is used to determine the spammer.

The results show that the proposed method is faster than any other method in large magnitude as dataset considered are very sparse. In my opinion, structural rank needs an additional attention measuring multi-document similarity, for example, evaluating document clusters, document relevance, or in search engines.

Author Fei et al. (2013) focused on detecting spammers that write reviews in short burst. They represented the reviewers- reviewers (reviews-reviews) relationship by graph. It used a graph propagation method to classify reviewers as spammers. The authors used Kernel Density Estimation techniques to develop an algorithm for detecting burst patterns in reviews for a specific product. The authors have also employed Markov random fields (MRFs) to model a reviewer in bursts. The authors relied on behavioral features to detect periods in time when review bursts per product coincided with reviewer burst. They used a hidden node to represent a reviewers real identity that might be spammer, non-spammer or mixed. The feature induced message passing in a loopy belief propagation framework is used to detect review spammers. The authors discarded singleton reviewers from the initial dataset, since they provided little behavior information. By discarding singleton reviewers, this method is similar to the one proposed by Mukherjee et al. (2012). This paper is only detected fake reviews written by elite users on a review platform. Authors claims, the proposed methods is better than Wang et al. (2011), because earlier method dose not identify the spammers. The weakness of this method seem that it misses the spammers detection in normal situations (when products are just launched). The spammers start their activity since the target product have been launched. Authors have formed the hypothesis for the review burst (only on spammers), but they have not considered the effect, when any product just become popular due to successful advertisement or any discounts given by the owner of products

Liang et al. (2014) have used the TrustRank method described in Gyöngyi et al. (2004) for spam detection. In TrustRank the semi-automatic technique is firstly used for classification of web pages into bad or good pages. Authors have represented each node as reviewer. The edge represented the inter-relationship between reviewers on products. Authors have introduced supportive edge and conflict edge concept. The supportive edge represented the same opinion on each product by pair of reviewer and conflict edge represented, different opinion on products. The unsupervised aspect computation framework has introduced, which is totally based on the reviewer unreliability in terms of score. This is the first algorithm in which authors emphasized on reviewer features and inter-relationship among the reviewers. The graph characteristics have integrated with reviewers features and spamming behavior. However, authors have not focused on the text or content of reviews. The reviews text is more useful and adds more significant results in spammer detection.

6 Discussion

This study presented many interesting and useful work regarding the state- of –art approaches in reviewer spammer detection. This paper is organized on the basis of techniques, reviewers features, datasets of products and reviews used in reviewers spammers detection.

If we considered group spammer features presented in Table 1, only five papers are published on detection of group spammers. Here, total numbers of 13 features are discovered by the researchers for model building. Most of group spammers features proposed by Mukherjee et al. (2012) in first group spammer detection paper. The detection of group spammer is complex task. First basic requirement for detection of review spammer is to construct the reviewer groups using dataset.

Mukherjee et al. (2012), Xu et al. (2013) have used frequent item-set mining algorithm for constructing groups of candidate reviewers. Choo et al. (2015) have used the user relationship graph for groups making. Here, the review text related groups features are more helpful to check the exact text matching. The groups spammer features are : Group Content Similarity(GCS), Group Member Content Similarity(GMCS) and Review Burstiness (BST). The size of spammer groups always plays vital roles to attract the total attention of other user. For normal users the size of groups are hidden because only through reading normal users cannot understand different person is encouraging each other reviews. The used features are: Group Size Ratio (GSR), Group Size (GS) and Group Support Count. As same way, the review posting time and days are very important factor of any products for boosting the sale of items. As same, the spammer groups tries to capture the possible attention at first glance by posting frauds reviews within limited time and day's frame. The time and days based features are: Group Time Window (GTW), Group Early Time Frame (GETF), Maximum One Day Review Ratio(MOR) and First Review Ratio. The rating of review is one of the factor which has used by many researchers to detect the review spammers (Lim et al. 2010; Liang et al. 2014). The spammer groups have given the ratings which is deviated by the normal user given rating. The features used for reviews rating are: Group Deviation (GD), Deviated Rating Ratio and Rating Abuse item Ratio(RA).

If we considered individual features based spammer detection, most of the individual features presented in Table 2. Here, total number of 42 features presented in Table 2 which is used in existing papers. Some of the features have commonly used by the researchers. Each review contain limited number of attributes, such as, review text, reviewer name and id, date and time of posting review, product details, rating of product given by reviewer and helpfulness feedback etc. To perform the spam detection on review datasets, we need to construct some new features from present set of attributes.

Some ratings features also used for detection of individual spammer. Researches have built some new features like, Individual Rating Deviation (IRD), General Deviation (GD), Average Standard Deviation (ASDR), Reviewer Rating Difference, Rating Deviation (RD)/ Rating Spamming/ Rating Abuse/ Rating Similarity, Single product Group Multiple High Rating/Low Rating, Extreme Rating (EXT) by the Reviewers and Average Rating.

The general tendency of spammer is, they give mostly two types of ratings to particular products, high or low ratings, here high means they gave 4 or 5 rating and low means 1 or 2 ratings. The Individual Rating Deviation (IRD), General Deviation (GD), Average Standard Deviation (ASDR) etc., are some example of rating deviation. The spammer gives false rating which is generally deviated from original user given rating. Sometimes spammer gives same rating or just gives average rating of products to create confusion among the users, for example, Rating Similarity and Average Rating. Rating similarity and average rating contains same characteristics but the definition and purpose of use is different with each other. The rating features are mostly used to model building by researches. All presented rating features mostly helps for individual spammer detection.

The time and date of the review posting is always very important for spammer detection. The spammer generally performed review spam activity, when a product is launched. The main purpose of spammers is to performed maximum damage or profit from posting spam reviews. When time and date has counted for review spammer detection than researchers have constructed some new features for example, Individual Early Time Frame (IETF), First Product Review, Minimum Time Interval, Early Time Frame, Maximum Number of Reviews (MNR) and Reviewing Burstiness (BST). The above time and date features researcher tries to capture the gesture activity of spammer.

The content of review is very top level features for spammer detection. The review text is the main part of review structure and has done maximum effect on human. The features are, Individual Content Similarity (ICS), Review Count (RC), Content Similarity (CS)/Review Similarity, Review Number, Review Text Spamming, Sentiment Score Based on Lexical Feature, Review Word Length Score, Duplicate/ Near Duplicate Reviews (DUP) and Length of Review used. Review spammers are in hurry to write spam review because of this reason most of time they just copied existing review or perform slightly modification in existing review. The detection of such type of changes has detected by some features like, Individual Content Similarity (ICS), Review Count (RC), Content Similarity (CS)/Review Similarity, Review Number and Review Text Spamming. As same, the spammers used sentiment of review by posting positive or negative reviews. The spammers have written positive reviews to increase the sale of products by attracting the positive mood of buyers and writes negative reviews to misguide other buyers by creating negative image of products. The length of review is one of the major features for detection of spammer. The short review has greater impact on buyer than lengthy reviews. The spammer mostly writes short reviews. Spammers have tried to post one review many time because if one is review visible many time, that create one personal space in human mind. Because of this reason, review count play vital role in this scenario. In same way other features describe in Table 2 play vital role in spammer detection.

Table 3 presents reviews dataset summary used by researchers. Here we have described the numbers of reviews, products and reviewers size used by the researches for detection purpose, irrespective of original numbers of reviews, products and reviewers present in datasets. The shown dataset mostly are from amazon e-commerce website because amazon is world largest e-commerce website in which millions of users used amazon services to buy goods. Other large amount of services provide by the tripadviosor.com. This is the world largest hotels online booking website. The size of the dataset, review size, reviewer size and store or products size are very huge because authors used large dataset to test accuracy of proposed model.

All dataset presented in table are not freely available. Some of the datasets has crawl by the authors. You can mail to authors to provide dataset for research purpose. There are some dataset like Amazon dataset (exclude M-Product dataset) is freely available.[5] This is quite normal before to use any dataset pre-processing is essential step. The pre-processing steps for review and reviewer have already explained in Section 3.

[5] http://snap.stanford.edu/data/web-Amazon-links.html

In Table 3 the Choo et al. (2015) have used the dataset of Amazon which contains book, movie, electronics, tools across datasets. Authors have used the maximum numbers of reviews, reviewers and products values for detection of spammers. The Amazon datasets have been repeatedly used by researchers because of the availability and size of the reviews, products and reviewers.

In Table 4 presented the comparison among existing published papers on the basis of Performance Metric, Kappa Static Results, Feature Used for Model Building, Results/Score In Percentage, Performance Validation Method Used and pros & cons components. The Performance metric basically explained the comparison done among existing method with proposed method. Most of the authors used rank normalized discounted cumulative gain (NDCG). Whereas some, authors have used accuracy, precision, f-1 measure and ROC metrics.

The Cohen Kappa or Fleiss Kappa are factual measure for evaluating the dependability of understanding between a settled number of raters when doing out clear cut appraisals to various things or grouping things. This stands out from different kappas, for example, Cohen's kappa, which is just worked while evaluating the assertion between two raters. Here, most of the authors have shown the human evaluation agreement results on Substantial Agreement.

In the next column present the "Feature Used for Model Building". Most of the authors have used the reviewer features, products features, review content features etc., for model building. In the next column present the results in percentage or scores. The performance validation of proposed model is a crucial task for checking the model accuracy with well-known techniques. Most of the authors have used human evaluation method for checking and comparing the model accuracy with human expert evaluation results. To perform human evaluation, the first steps to select experts from e-commerce industry, who have an experience in tagging of reviews in spam or non-spam. In next step we set some guidelines for human experts. This is suited for comparing proposed model. The guidelines should not be more than five or six points. Using guidelines experts marked the review as a spam or non-spam. Some of the authors have also used rank based method for validation. There are supervised and unsupervised algorithms which also help for performance validation of model. In Table 5, we have presented the pros and cons of existing papers. Here, we have seen that the area of group spammer detection is less explored and required much more efforts for taking the challenges it offers.

6.1 Open problems and limitations

While summarizing the state-of-art approaches in review spammer detection, we observed number of open issues and challenges. Some are:

6.1.1 Paucity of labeled review datasets

The biggest problem in review spam and spammers detection is availability of label data. One labeled data of hotel reviews is available. It has been made by Ott et al. (2011). This dataset have some limitations, such as limitation in number of attributes. We need labeling of data to train a classifier which will then classify a unknown review as a spam or non-spam. For labeling of dataset, researchers are dependent on the manual techniques (human judges). Ott et al. (2011) have used AMT (Amazon Mechanical Trunk) for producing the label dataset for spam detection, but Mukherjee et al. (2013) suggested that proposed method of labeling has not given better results for review spam detection in real dataset.

Limited number of work is done on real world dataset. In current scenario human experts are widely used in detection of the spam reviews. This method is very time consuming, and has constrained accuracy which is susceptible to failure on larger and high speed data. Most of the researchers have used Amazon.com dataset for experiment purpose, but data labeling problem exists in their dataset as well. Label dataset problem could be solved by introducing new label datasets or by labeling existing dataset. Already labeled review dataset that are used by different researchers could be made openly available for further research purpose.

6.1.2 Rapidly growing volume of review dataset

Data collection and handling is major challenge of review spam detection. Though there are various crawlers freely available, that can be used but each has its own constraints. The dimensions of the data crawled could be less which may affect quality. Data may be better procured. The actual reviews can be very huge, and coming with high velocity. There are million of reviews and reviewers and are increasing day by day. Like currently Amazon.com are having 244 million active users and millions of reviews and reviewers. To perform experimentation on large dataset requires high computing power. The reviews may also be in multilingual form and may even contain different dialects. Review text length is also not fixed. It can be one or two lines and one or two pages processing. As we know that data mining need maximum effort (50 % to 60 %) in pre-processing of data (Tan et al. 2006). For pre-processing of the data some open source tools that support English text is used, but for processing reviews in Chinese, Hindi or local dialects new tools are required, thus enriching text pre-processing. New feature extraction on large dataset is computationally time taking process. One of the main open challenge in review spam and spammer is the use of semantic algorithms. There is less work done on semantic based algorithms. The Semantic analysis of words depends upon the WordNet and SentiWordNet. The WordNet and SentiWordNet contains big dictionary of words which is used for sentiment as well as semantic analysis of reviews. Up-till now, the semantic based model is not proposed for review group spammer detection.

6.1.3 Limited numbers of attributes and unavailability of major attributes

The present attributes in review dataset are limited. The available attributes are not able to detect review spams accurately. Availability of multi-dimensional review datasets is one of major issue. Researchers depend upon various crawlers to crawl review dataset or use already crawled dataset reported by other researchers. The crawled datasets contains limited number of attributes. For example, in Amazon dataset there are only ten attributes presents which are following:
(product/productId),(product/title),(product/price),
(review/userId),(review/profileName),
(review/helpfulness),(review/score),
(review/time)(review/summary),(review/text) (McAuley and Leskovec 2013). To increase the review spam detection accuracy, researchers require more attributes, for example, reviewer registered email-id, review posting location of reviewer, IP address of system from which review is posted etc., but these attributes are not openly available on company websites. The location of reviewer feature is used in Sandulescu and Ester (2015). E-commerce companies could be providing above features for review spam detection purpose, while they may ask for singing a legal contract or accepting to certain terms & conditions to stop misuse of dataset.

Ma and Li (2012) have suggested more attributes for detection of review spam and spammer which are following:

Associate review credibility with verifiable actions (Ma and Li 2012) There are few e-commerce companies which use one more features to check reviewer credibility for example, By Certified Buyers, Verified Purchase etc. The "*By Certified Buyers*" feature is used by Filpkart Company which shows that the review is written by actual buyer. In same manner, Amazon is providing the "Verified Purchase" Review Service for credibility checking and third-party review sites. It is provided as the "certified review" where the reviewers mostly visited the site. These reviews may be considered as genuine reviews and reviewers may be considered as genuine reviewers. Review credibility is increased by including some more attributes; such as the name of product purchased, date & time of purchase or services used date & time. In case of reviewer profile openly available to read and all reviews written by that reviewer are linked with profile. These features may be further help in review spam detection.

Associate review credibility with location information (Ma and Li 2012) The location information of posting review could be utilized to assess the credibility of reviews submitted by reviewer using social networking services like, Foursquare or GetGlue that allow users to "check in" to a physical place of consumption with some incentives. Locations (country and city) of users are also tracked by the IP address of system or tracking the current location of mobile phone when reviewer is posting review.

Associate review credibility with social relationships (Ma and Li 2012) At the time of account opening by users for writing reviews, there must be provision of linking of users social media account with opened account. E-commerce companies can put a compulsion for linking of any valid id-proof for account opening. The researchers are using the social relationships of users for detection review spams. The relationships of reviewer with his followers utilized to measure the credibility with one's posting review to his connected friends or followers. The "helpful" feature is used by E-commerce websites to check how many users in favor or in against of review. Spammer is also performed spam activity on "helpful" feature by using some automate program which increases or decreases counts of "helpful" feature. This type of spam activity used spammer can be stop by using captcha which is linked with the review helpful feedback click.

Need of attributes availability As we have discussed that review datasets are crawled from the E-commerce web-sites. Crawled reviews contain limited attributes. Review spams and spammers detection are performed in more accurate way if some more attributes are available like; IP address, registered email-Id for sign-up, location at first login by IP address, product purchased date & time or services used date & time, reviewer profile are openly available to read and reviewer written all reviews are linked with profile, reviewer profile linked with social media profile and linking of any valid id-proof for account opening etc. All above features are further used for review spam and spammer detection model building, which can enrich the detection accuracy.

7 Future works

The research in review spammer detection just started since 2010–2011. A very limited work is done by researchers which is based on some basic techniques. There are plenty of improvement and unexplored area for future research works. We have given some possible future directions which could be used in future research.

7.1 Integration of different domain dataset for more accurate analysis

Pattern analysis of review spam and spammer are helpful to understand the behaviour and writing of spammer. The integration of different review data, followed by patterns analysis on integrated data is an unexplored area of research. The pattern analysis of spammer could be performed by combined study of different review dataset for example; Amazon reviews dataset and Filpkart review dataset could integrated for group spammer behaviour analysis. There are some typical patterns in which spammers interact with each other in different domain dataset. Detection of such patterns and then exploring this information for detecting spammers has huge potential.

7.2 Tagging of reviews as spam/non-spam and spammer/non-spammer in real time streaming of reviews

The velocity of writing of review is increasing day by day. All earlier research work in review spam and spammer detection is based on the static dataset. There is limited work done reported on tagging of review spam and review spammer on static dataset (Mukherjee et al. 2012). To find out review spam and spammer in static dataset is time taking process, due to this reason spammer gain maximum advantage by posting fake reviews when product is just lunched. Spam and spammer tagging in real time streaming of reviews are not explored by researchers.

7.3 Identifying review spammer communities by analysis of reviewers profiles and comments or feedbacks on reviews

There is less work done by researchers in detection of review group spammer. All earlier works are based on the review datasets and behaviour of reviewers. The profile of reviewer is not used by the researcher for detection of review spammer communities. In same manner, there are some followers of posted reviews in which followers also show their interest or experiences on posted review by giving their own view in written form. The comments or feedbacks of reviews are not taken as feature for detection of review spam and spammer by researchers which use in future.

7.4 Multilingual review spam detection

The review spam detection is now switched over to multilingual language analysis. As review is user generated content which enables user to write in their own language and dialect. Some researchers have worked on different language dataset like using English, Chinese, Arabic etc., but there is a need to analyze multilingual reviews. As we know, spammer are in hurry to write, they most of time copy exact text from other dataset. Spammer may also use some language conversion tool (like Google Translate) to convert the Chinese review into English or any other languages, so there is need for the analyses of such type of spam and spammer detection in future.

8 Conclusion

This survey paper gives an overview of the various spammer detection techniques that has been used till date. The main idea of this survey is to furnish the state of art in review spammer detection to professionals and analysts and future directions. It is observed that, basically two types of spammer are present, group spammer and individual spammer. The detection of these two types of spammers is different from each other. Only thirteen features have been proposed till date for group spammer detection. This survey is give notion that the earlier research is mainly focused on individual spammer detection. This survey may help the researchers to find the effective methods and features for review spammer detection. Researchers can use these features in their research. Besides this, it can assist arbitrators of different e-commerce company's websites with providing proper strategies for their site and enabling new guidelines for account opening which are discussed in this survey. Even users of opinion websites can get benefit from this survey and they can get a broad view about the spammers. The future directions of research in review spammers detection is directed towards the use of hybrid and unsupervised techniques. The spammers detection could be more accurately performed, if more attributes (as listed in Section 6.1) are available.

References

Akoglu, L., Chandy, R., & Faloutsos, C. (2013). Opinion fraud detection in online reviews by network effects. *ICWSM*, *13*, 2–11.

Aye, C.M., & Oo, K.M. (2014). Review spammer detection by using behaviors based scoring methods. In *Proceedings of international conference on advances in engineering and technology*.

Baeza-Yates, R.A., Castillo, C., López, V., & Telefónica, C. (2005). Pagerank increase under different collusion topologies. In *AIRWeb* (Vol. 5, pp. 25–32).

Berger, P., Hennig, P., Schoenberg, M., & Meinel, C. (2015). Blog, forum or newspaper? Web genre detection using svms. In *2015 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology (WI-IAT)* (pp. 364–68). IEEE.

Carpinter, J., & Hunt, R. (2006). Tightening the net: a review of current and next generation spam filtering tools. *Computers & Security*, *25*(8), 566–578.

Choo, E., Yu, T., & Chi, M. (2015). Detecting opinion spammer groups through community discovery and sentiment analysis. In *Data and applications security and privacy XXIX* (pp. 170–187). Springer.

Choudhury, S., Dey, B., & Kumar, S. (2005). Spam: a threat to network security in digital library and information centres.

Crawford, M., Khoshgoftaar, T.M., Prusa, J.D., Richter, A.N., & Al Najada, H. (2015). Survey of review spam detection using machine learning techniques. *Journal of Big Data*, *2*(1), 1–24.

Duh, A., Štiglic, G., & Korošak, D. (2013). Enhancing identification of opinion spammer groups. In *Proceedings of international conference on making sense of converging media* (Vol. 326). ACM.

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: a publicly available lexical resource for opinion mining. In *Proceedings of LREC* (Vol. 6, pp. 417–422). Citeseer.

Fayazbakhsh, S.K., & Sinha, J. (2012). Review spam detection: a network-based approach. Final Project Report: CSE, 590.

Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Exploiting burstiness in reviews for review spammer detection. *ICWSM*, *13*, 175–184.

Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378.

Gyongyi, Z., & Garcia-Molina, H. (2005). Web spam taxonomy. In *First international workshop on adversarial information retrieval on the web (AIRWeb 2005)*.

Gyöngyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). Combating web spam with trustrank. In *Proceedings of the thirtieth international conference on very large data bases - volume 30, VLDB '04* (pp. 576–587). VLDB Endowment.

Heydari, A., Ali Tavakoli, M., Salim, N., & Heydari, Z. (2015). Detection of review spam: a survey. *Expert Systems with Applications*, *42*(7), 3634–3642.

Hinde, S. (2002). Spam, scams, chains, hoaxes and other junk mail. *Computers & Security*, *21*(7), 592–606.

Huang, J., Qian, T., He, G., Zhong, M., & Peng, Q. (2013). Detecting professional spam reviewers. In *Advanced data mining and applications* (pp. 288–299). Springer.

Jindal, N., & Liu, B. (2007). Analyzing and detecting review spam. In *Seventh IEEE international conference on data mining, 2007. ICDM 2007* (pp. 547–552). IEEE.

Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the 2008 international conference on web search and data mining, WSDM '08* (pp. 219–230). ACM, New York.

Jiang, B., Chen, B., et al. (2013). Detecting product review spammers using activity model. In *2013 international conference on advanced computer science and electronics information (ICACSEI 2013)*. Atlantis Press.

Kim, S., Park, H., & Lebanon, G. (2014). Fast spammer detection using structural rank. arXiv:1407.7072.

Li, W., Zhong, N., & Liu, C. (2006). Combining multiple email filters based on multivariate statistical analysis. In *Foundations of intelligent systems* (pp. 729–738). Springer.

Liang, D., Liu, X., & Shen, H. (2014). Detecting spam reviewers by combing reviewer feature and relationship. In *2014 international conference on informative and cybernetics for computational social systems (ICCSS)* (pp. 102–107). IEEE.

Lim, E.-P., Nguyen, V.-A., Jindal, N., Liu, B., & Lauw, H.W. (2010). Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on information and knowledge management, CIKM '10* (pp. 939–948). ACM: New York, NY, USA.

Lu, Y., Zhang, L., Xiao, Y., & Li, Y. (2013). Simultaneously detecting fake reviews and review spammers using factor graph model. In *Proceedings of the 5th annual ACM web science conference* (pp. 225–233). ACM.

Luckner, M., Gad, M., & Sobkowiak, P. (2014). Stable web spam detection using features based on lexical items. *Computers & Security*, *46*, 79–93.

Ma, Y., & Li, F. (2012). Detecting review spam: challenges and opportunities. In *2012 8th international conference on collaborative computing: networking, applications and worksharing (CollaborateCom)* (pp. 651–654). IEEE.

McAuley, J., & Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on recommender systems* (pp. 165–172). ACM.

Mukherjee, A., Liu, B., Wang, J., Glance, N., & Jindal, N. (2011). Detecting group review spam. In *Proceedings of the 20th international conference companion on World wide web* (pp. 93–94). ACM.

Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web, WWW '12* (pp. 191–200). ACM: New York.

Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Spotting opinion spammers using behavioral footprints. In *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 632–640). ACM.

Ott, M., Choi, Y., Cardie, C., & Hancock, J.T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies-volume 1* (pp. 309–319). Association for Computational Linguistics.

Peng, Q. (2014). Store review spammer detection based on review relationship. In *Advances in conceptual modeling* (pp. 287–298). Springer.

Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A bayesian approach to filtering junk e-mail. In *Learning for text categorization: papers from the 1998 workshop* (Vol. 62, pp. 98–105).

Sandulescu, V., & Ester, M. (2015). Detecting singleton review spammers using semantic similarity. In *Proceedings of the 24th international conference on World Wide Web* (pp. 971–976). ACM.

Seneviratne, S., Seneviratne, A., Kaafar, M.A., Mahanti, A., & Mohapatra, P. (2015). Early detection of spam mobile apps. In *Proceedings of the 24th international conference on World Wide Web, WWW '15* (pp. 949–959). ACM, New York.

Tan, P.-N. et al. (2006). *Introduction to data mining*. India: Pearson Education.

Vorakulpipat, C., Visoottiviseth, V., & Siwamogsatham, S. (2012). Polite sender: a resource-saving spam email countermeasure based on sender responsibilities and recipient justifications. *Computers & Security*, *31*(3), 286–298.

Wang, J., & Liang, X. (2013). Discovering the rating pattern of online reviewers through data coclustering. In *2013 IEEE international conference on intelligence and security informatics (ISI)* (pp. 374–376). IEEE.

Wang, G., Xie, S., Liu, B., & Yu, P.S. (2011). Review graph based online store review spammer detection. In *Proceedings of the 2011 IEEE 11th international conference on data mining, ICDM'11* (pp. 1242–1247). IEEE Computer Society: Washington, DC, USA.

Wang, G., Xie, S., Liu, B., & Yu, P.S. (2012). Identify online store review spammers via social review graph. *ACM Transactions on Intelligent Systems and Technology*, *3*(4), 61:1–61:21.

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). Opinionfinder: A system for subjectivity analysis. In *Proceedings of hlt/emnlp on interactive demonstrations* (pp. 34–35). Association for Computational Linguistics.

Wu, B., & Davison, B.D. (2005). Identifying link farm spam pages. In *Special interest tracks and posters of the 14th international conference on World Wide Web, WWW '05* (pp. 820–829). ACM, New York.

Wu, B., Goel, V., & Davison, B.D. (2006). Topical trustrank: Using topicality to combat web spam. In *Proceedings of the 15th international conference on World Wide Web, WWW '06* (pp. 63–72). ACM, New York.

Xu, C., Zhang, J., Chang, K., & Long, C. (2013). Uncovering collusive spammers in chinese review websites. In *Proceedings of the 22nd ACM international conference on conference on information & knowledge management* (pp. 979–988). ACM.

Yuan, G.-X., Ho, C.-H., & Lin, C.-J. (2012). Recent advances of large-scale linear classification. *Proceedings of the IEEE*, *100*(9), 2584–2603.

Zhou, Y. (2011). Structure learning of probabilistic graphical models: a comprehensive survey. arXiv:1111.6925.