Evidential Group Spammers Detection

Malika Ben Khalifa, Zied Elouedi, Eric Lefevre

Evidential Group Spammers Detection

Malika Ben Khalifa[1,2], Zied Elouedi[1], and Eric Lefèvre[2]

[1] Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisia
malikabenkhalifa2@gmail.com, zied.elouedi@gmx.fr
[2] Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Béthune, F-62400, France
eric.lefevre@univ-artois.fr

Abstract. Online reviews are considered as one of the most prevalent reference indicators for people to evaluate the quality of different products or services before purchasing. Since these reviews affect the buying decision of customers and control the success of the different e-commerce websites, the activity of fake reviews posting is more and more increasing. These fraudulent reviews are posted by a large number of spammers who try to promote or demote target products or companies. The reviewers spammers generally work collaboratively under group of spammers to take control of reviews given to some products, which seriously damage the review system. To deal with this issue, we propose a novel method aim to detect group spammers while relying on various group spamming behavioral indicators. Our approach is based on the K-nearest neighbors algorithm under the belief function theory to treat the uncertainty in the used behavioral indicators. Our method succeeds in distinguishing between genuine and fraudulent group of reviewers. It was tested on two large real datasets extracted from yelp.com.

Keywords: Fake reviews, Group spammers, Uncertainty, Belief Function Theory, Evidential KNN, E-commerce.

1 Introduction

Products, brands, hotels, restaurant, cities, places to visit and all services are now identified through a rating score which is generally the average score of the different reviews given by customers. Such rating score or reviews become one of the most influenced source on consumer's purchase decisions. We can assume that, online reviews nowadays control e-commerce and even international commerce. To increase their market share and to stay ahead of their competitors, companies and business try to over qualify their products by posting fake positive reviews, and even by posting fake negative reviews to damage their competitors' e-reputation. Those who post these fake reviews are called fake reviewers or review spammers, and the products being spammed are called target products. As the commercialization of these fraudulent activities, such spammers are organized to collaboratively write fake reviews in order magnify the effect of review manipulation. Such review group spammers are more frequently occurred

to control the sentiment of the target products. They are even more harmful than individual review spammers' cause over their ability to deviate the overall rating in a short time interval and with different reviewers profiles to mislead spammer detection tools. The review spam detection issue attracts significant researchers during the last years. The main objective of these methods is to distinguish between fake and genuine reviews in order to protect ensure a safe environment and an equitable concurrence between companies. These research can be classified into three categories [11]; review spam detection, review spammer detection and group spammer detection. Several approaches are based on the review spam text information as the semantic and linguistic aspects [7, 17]. Moreover, there are different methods which try to detect spammers through graph based aspects [1, 8, 23]. Others, detect spammers while relying on the spammers behavioral indicators [12, 15, 18], and on the brust patterns as new indicators [9]. These approaches give significant results in the spam reviews detection field. Recently, there were increasingly research interests in group spamming detection aspects cause of their powerful manipulation thanks to their huge reviewers' members. The first study was introduced by Mukherjee et al. [14], in which they rely on the Frequent Itemset Mining (FIM) technique to generate candidate review spammers groups. This technique considers reviewers as items and products as transactions. Through the FIM technique and by initializing the minimum support count to 3, they can spot at least 2 reviewers, while each reviewer review at least 3 common products. Many techniques rely on these candidate groups and propose different computing frameworks to evaluate the suspicion of each candidate spammer groups. Such that, in [14] authors proposed an iterative computing GSRank to rank candidate groups which spots the relationship among candidate groups, target products and individual reviewers. Xu et al. [27] introduce a statistical model based on the EM algorithm to calculate the collusiveness of each group member from one FIM candidate group at least. Another proposed method in [24] relies on FIM method to capture bicliques or sub-bicliques candidates them check them to detect real collusion groups through group spam indicators. Moreover, Xu et al. [28] use FIM to find groups of reviewers who have reviewed various common products. They introduce a KNN-based method and a graph based classification method to predict the fake or not fake labels for each reviewer belonging to at least one FIM candidate group. They evaluated the effectiveness of the used group spammer indicators on a large Chinese review websites. The KNN-method proves its performance in this study. Some other recent works in this aspect do not rely on the FIM techniques such in [25], where authors propose a top-down computing framework (GGSpam) to detect review spammer groups by exploiting the topological structure of the underlying reviewer graph. We can also cite [26] which propose an unsupervised approach named LDA-based group spamming detection in product reviews (GSLDA) which adapt Latent Dirichlet Allocation (LDA) in order to bound the closely related group spammers into small cluster of reviewers and extracts high suspicious reviewers groups from each LDA-clusters. These proposed methods achieve also significant results. The spam review detection issue can be considerate as one of the most uncertain challenging problem

due to the ambiguity provided by the spammers and the group spammers to mislead the detection systems. Nevertheless, the previous proposed methods did not take into consideration the uncertain aspect while trying to detect group spammers. We think that ignoring such uncertainty may deeply affect the quality of detection. For these reasons, we propose a novel method aims to detect group spammers based on the FIM technique to generate candidate group and also on the different group spammer indicators, using the K-nearest neighbors' algorithm within the belief function theory. This theory has shown its robustness in this field through our previous methods which achieve significant results [4, 3, 2]. Furthermore, the use of the Evidential K-NN has been based on its robustness in the real world classification problems under uncertainty. We seek to involve imprecision in the Group spammers behaviors indicators which are considered as the fundamental interest in our approach since they are used as features for the Evidential K-NN. In such way, our method predicts the labels spammers or not spammers reviewers (belonging or not to the FIM candidate groups).
This paper is structured as follows: In the first section, we present the basic concepts of the belief function theory and the Evidential K-nearest neighbors, then we elucidate the proposed method in section 2. Section 3 is consecrated for the experimental results and we finish with a conclusion and some future work.

2 Belief function theory

In section, we elucidate the fundamentals of the belief function theory as well as the Evidential K-nearest neighbors classifier.

2.1 Basic concepts

The belief function theory, called also the Dempster Shafer theory, is one of the powerful theories that handles uncertainty in different tasks. It was introduced by Shafer [20] as a model to manage beliefs.
In this theory, a given problem is represented by a finite and exhaustive set of different events called the frame of discernment Ω. 2^Ω is the power set of Ω that includes all possible hypotheses and it is defined by: $2^\Omega = \{A : A \subseteq \Omega\}$.
A basic belief assignment (*bba*) named also a belief mass represents the degree of belief given to an element A. It is defined as a function m^Ω from 2^Ω to $[0, 1]$ such that:

$$\sum_{A \subseteq \Omega} m^\Omega(A) = 1. \tag{1}$$

A focal element A is a set of hypotheses with positive mass value $m^\Omega(A) > 0$.

Several types of *bba* 's have been proposed [21] in order to model special situations of uncertainty. Here, we present some special cases of *bba*'s:

- The certain *bba* represents the state of total certainty and it is defined as follows: $m^\Omega(\{\omega_i\}) = 1$ and $\omega_i \in \Omega$.

- The categorical *bba* has a unique focal element A different from the frame of discernment defined by: $m^{\Omega}(A) = 1$, $\forall A \subset \Omega$ and $m^{\Omega}(B) = 0$, $\forall B \subseteq \Omega$ $B \neq A$.
- Simple support function: In this case, the *bba* focal elements are $\{A, \Omega\}$. A simple support function is defined as the following equation:

$$m^{\Omega}(X) = \begin{cases} w & \text{if } X = \Omega \\ 1 - w & \text{if } X = A \text{ for some } A \subset \Omega \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where A is the focus and $w \in [0,1]$.

Belief function

The belief function, denoted *bel*, includes all the basic belief masses given to the subsets of A. It quantifies the total belief committed to an event A by assigning to every subset A of Ω the sum of belief masses committed to every subset of A. *bel* is represented as follows:

$$bel(A) = \sum_{\emptyset \neq B \subseteq \Omega} m^{\Omega}(B) \tag{3}$$

$$bel(\emptyset) = 0 \tag{4}$$

Combination Rules

Several combination rules have been proposed in the framework of belief functions to aggregate a set of *bba*'s provided by pieces for evidence from different experts. Let m_1^{Ω} and m_2^{Ω} two *bba*'s modeling two distinct sources of information defined on the same frame of discernment Ω. In what follows, we elucidate the combination rules related to our approach.

1. *Conjunctive rule*: It was introduces in [22], denoted by $\textcircled{\cap}$ and defined as:

$$m_1^{\Omega} \textcircled{\cap} m_2^{\Omega}(A) = \sum_{B \cap C = A} m_1^{\Omega}(B) m_2^{\Omega}(C) \tag{5}$$

2. *Dempster's rule of combination*: This combination rule is a normalized version of the conjunctive rule [5]. It is denoted by $\oplus$ and defined as:

$$m_1^{\Omega} \oplus m_2^{\Omega}(A) = \begin{cases} \frac{m_1^{\Omega} \textcircled{\cap} m_2^{\Omega}(A)}{1 - m_1^{\Omega} \textcircled{\cap} m_2^{\Omega}(\emptyset)} & \text{if} \quad A \neq \emptyset, \forall A \subseteq \Omega, \\ 0 & otherwise. \end{cases} \tag{6}$$

Decision process

The belief function framework provides numerous solutions to make decision. Within the Transferable Belief Model (TBM) [22], the decision process is

performed at the pignistic level where *bba's* are transformed into the pignistic probabilities denoted by $BetP$ and defined as:

$$BetP(B) = \sum_{A \subseteq \Omega} \frac{|A \cap B|}{|A|} \frac{m^{\Omega}(A)}{(1 - m^{\Omega}(\emptyset))} \quad \forall\ B \in \Omega \tag{7}$$

2.2 Evidential K-Nearest neighbors

The Evidential K-Nearest Neighbors (EKNN) [6] is one of the best known classification methods based in the belief function framework. It performs the classification over the basic crisp KNN method thanks to its ability to offer a credal classification of the different objects. This credal partition provides a richer information content of the classifier's output.
Notations

- $\Omega = \{C_1, C_2, ..., C_N\}$: The frame of discernment containing the N possible classes of the problem.
- $X_i = \{X_1, X_2, ..., X_m\}$: The object X_i belonging to the set of m distinct instances in the problem.
- A new instance X to be classified.
- $N_K(X)$: The set of the K-Nearest Neighbors of X.

EKNN method
The main objective of the EKNN is to classify a new object X based on the information given by the training set. A new instance X to be classified must be allocated to one class of the $N_K(X)$ founded on the selected neighbors. Nevertheless, the knowledge that a neighbor X_i belongs to class C_q may be deemed $d(X, X_i)$ as a piece of evidence that raises the belief that the object X to be classified belongs to the class C_q. For this reason, the EKNN technique deals with this fact and treats each neighbor as a piece of evidence that support some hypotheses about the class of the pattern X to be classified. In fact, the more the distance between X and X_i is reduces, the more the evidence is strong. This evidence can be illustrated by a simple support function with a *bba* such that:

$$m_{X,X_i}(\{C_q\}) = \alpha_0 \exp^{-(\gamma_q^2 d(X,X_i)^2)} \tag{8}$$

$$m_{X,X_i}(\Omega) = 1 - \alpha_0 \exp^{-(\gamma_q^2 d(X,X_i)^2)} \tag{9}$$

Where:

- α_0 is a constant that has been fixed in 0.95.
- $d(X, X_i)$ represents the Euclidean distance between the instance to be classified and the other instances in the training set.
- γ_q assigned to each class C_q has been defined as a positive parameter. It represents the inverse of the mean distance between all the training instances belonging to the class C_q.

After the generation of the different $bba's$ by the K-nearest neighbors, they can be combined through the Dempster combination rule as follows:

$$m_X = m_{X,X_1} \oplus ... \oplus m_{X,X_K} \tag{10}$$

where $\{1, ..., K\}$ is the set including the indexes of the K-Nearest Neighbors.

3 Proposed Method

The idea behind our method is to take into account the uncertain aspect in order to improve detecting the group spammer reviewers. For that, we propose a novel approach based on FIM techniques to generate candidate groups, different group spammers indicators and we rely on the Evidential K-nearest neighbors which is famous classifier under the belief function framework. In the remainder of this section, we will elucidate the different steps of our proposed approach; in the first step we will construct the different group spammers from data and we model the group spammers indicators which will be used as features in our method. In the second step, we detail the applying of the EKNN in which we present the initialization and learning phase. Finally, we distinguish between the group spammers and the innocent reviewers through the classification phase.

3.1 Step1: Pre-processing phase

Spammers who get paid to post fake reviews can not just writing one review for a single product because they would not make enough money that way. Rather, they post various reviews for many products. That's why, we use Frequent pattern mining, that can find them working together on multiple products, to construct candidate spammer groups. Then, we elucidate the different group spammers indicators [14] which can control the candidate spammers behaviors and to find out whether these groups behave strangely.

1- Construction of spammer groups from data

To create a dataset that holds sufficient colluders for evaluation, the first task is to search for the places where colluders would probably be found. A good way to achieve this is to use frequent itemset mining (FIM). In such context, reviewer IDs are regarded as items, each transaction is the set of reviewer IDs who have reviewed a particular product.
Through FIM, groups of reviewers who have reviewed multiple common products can be found. Here we use maximal frequent itemset mining (MFIM) to discover groups with maximal size since we focus on the worst spamming activities in our dataset.

2- Group spammer indicators

In our method, we rely on these different group spammers indicators:

- Time Window (TW): Reviewers in a spammer group usually work together in order to post fake reviews for a target product in a short time interval.
- Group Deviation (GD): Members of group spammers are generally give either very high (5*) or very low (1*) ratings to the products. The same products typically are also reviewed by other genuine reviewers. Group spammers generally deviate in their ratings by a significant amount from the mean review ratings score. Therefore, the bigger the deviation, the worse the group is.
- Group Content Similarity (GCS): Members of group spammers usually copy reviews among themselves. Therefore, the products or the services which are victims of such group spamming can have many reviews with similar content.
- Member Content Similarity (MCS): The members of a group may not know one another. Each of them just copy or modify his/her own previous reviews. If multiple members of the group do this, the group is more likely to be a spammer group.
- Early Time Frame (ETF): One damaging group spam activity is to strike right after a product is launched or is made available for reviewing. The purpose is to make a big impact and to take control of the sentiment on the product.
- Ratio of Group Size (RGS): The ratio of the group size and the total number of reviewers for the product is also a good indicator of spamming. In one extreme (the worst case), the group members are the only reviewers of the product, which is very damaging.
- Group Size (GS): The group size itself also tells something quite interesting. If a group is large, then the probability of members happening to be in the group by chance is small. Furthermore, the larger the group, the more devastating is its effect.
- Support Count (SC): Support count is the number of products for which the group has worked on together. If a group has a very high support count, it is clearly alarming.

3.2 Step2: Evidential KNN application

After applying the FIM algorithm with fixed parameter settings, we note that the suspicious groups are found to be highly similar with each other in term of members, reviewed product also similar ratings. This is because of the dense reviewer product bipartite graph that can decomposed into many small pieces of fully connected sub-graphs (groups). This may have many overlapped nodes (members and products) and such small sub-graphs may dilute the effectiveness of some indicators. However, being used properly, these tiny groups may be favorable to detect colluders in a novel way.
That's why, we propose to rely on the Evidential KNN-based method to detect colluders by utilizing the similarities between such groups. Let $\{g_j\}$ $j = 1..m$ a set of groups and $\{R_i\}$ $i = 1..n$ be a set of reviewers with each associated with an vector a_i of attributes which are the different group spammer indicators mentioned above. Note that each reviewer may belong to multiple groups.

By modeling the colluder detection problem as a binary classification problem our goal is to assign each reviewer R_i with a class label $\Omega = \{S, \bar{S}\}$ where S represents the class of the spammers reviewers and $\bar{S}$ contains the class of the not spammers (innocent) reviewers.
The idea is that given a set of groups, the reviewers who belong to "similar" groups may be more likely to have the same class labels. Thus the class label of a reviewer R_i can be determined commonly by a set of k reviewers who belong to groups most "similar" to the groups R_i belongs to.

1-Initialization and learning phase

When applying the Evidential K-NN classifier, we start by initializing the parameters α_0 et γ_0 to be used in the learning phase. The α_0 is fixed to 0.95, as mentionned in the EKNN algorithm [6]. To ensure the γ_{Ii} computation performance, first of all we must find reviewers belonging to different groups are having separately exclusive group spammers indicators. We measure the pairwise similarity of two groups which consists of three measurements as follows:
Common Member Ratio
It measures the Jaccard similarity of the sets of members of two groups:

$$S_{cm} = \frac{|M_i \cap M_j|}{|M_i \cup M_j|} \tag{11}$$

Where M_i and M_j are the member sets of groups g_i and g_j.
Common Product Ratio
It is computed as the sum of the number of products (hotel/ restaurant) of the same brand reviewed by each group, divided by the sum of the number of products reviewed by each group:

$$S_{cp} = max_{b \in B} \frac{(P_{b,i}) + (P_{b,j})}{P_i + P_j} \tag{12}$$

where B is the set of common brands reviewed by both groups g_i and g_j. $P_{b,i}$ (respectively $P_{b,j}$) is the set of the products with brand b reviewed by group g_i (respectively g_j), and P_i (respectively P_j) is the set of the products reviewed by group g_i (respectively g_j).
Common Rating Deviation
It computes the deviation between the average ratings given to the products of common restaurant/hotel reviewed by two groups:

$$S_{crd} = \frac{1}{1 + \sqrt{\frac{1}{|B|} \sum_{b \in B} (\bar{r}_{b,i} - \bar{r}_{b,j})^2}} \tag{13}$$

where $\bar{r}_{b,i}$ (respectively $\bar{r}_{b,j}$) is the average rating given to the products with brand b by group g_i (respectively g_j). Accordingly, the pairwise similarity of two groups is defined as the weighted average of the above components:

$$Sg_{i,j} = \frac{w_k S_k}{\sum w_k S_k} \tag{14}$$

where $S_k \in \{S_{cp}, S_{crd}, S_{cm}\}$ and w_k is a non negative weight for S_k where $\sum_k w_k = 1$.
After defining the pairwise similarity of two groups, the pairwise similarity of two reviewers is computed by taking the average over the pairwise similarity of each pair of their respective groups:

$$d(R_i, R_j) = \frac{\sum_{k \in G_i} \sum_{l \in G_j} Sg_{k,l}}{|G_i||G_j|} \tag{15}$$

where G_i and G_j are the set of groups that have reviewer R_i and R_j respectively. Then, we must select a set of reviewers and for each reviewer R_j in the database, we measure its distance with the target reviewer R_i. Given a target reviewer, we have to select its K-most similar neighbors, by choosing only the K reviewers having the smallest distances values that is calculated through the pairwise similarity of two reviewers calculated above.

2- Classification phase

In this part, we aim to classify the target reviewer R_i into spammer or not where our frame of discernment $\Omega = \{S, \bar{S}\}$.

The *bba*'s generation

Each reviewer R_I provides a piece of evidence that represents our belief about the class that he belongs. However, this information does not offer certain knowledge about the class. In the belief function framework, this case is represented by simple support functions, where only a part of belief is assigned to $\omega_i \in \Omega$ and the rest is committed to Ω. Consequently, we obtain *bba* as follows:

$$m_{R_i,R_j}(\{\omega_i\}) = \alpha_{R_i} \tag{16}$$

$$m_{R_i,R_j}(\Omega) = 1 - \alpha_{R_i} \tag{17}$$

Where R_i is the target reviewer and R_j is its similar reviewer that $j = \{1..K\}$, $\alpha_{R_i} = \alpha_0 \exp^{(-\gamma_{Ii} d(R_i,R_j))}$, α_0 and γ_{Ii} are two parameters and $d(R_i, R_j)$ is the distance between the two reviewers R_i and R_j measured above.
In our situation, each neighbor of the target reviewer has two possible hypotheses. It can be near to a spammer reviewer in which his the committed belief is assigned to the spammer class S and the rest is given Ω. On the contrary, it can be similar to an innocent reviewer where the committed belief is allocated to the not spammer class $\bar{S}$ and the rest to the whole frame Ω. We treat the K-most similar reviewers independently where each one is represented by a *bba*. Hence, K various *bba*'s can be created for each reviewer.

The *bba*'s combination

After the *bba*'s generation for each reviewer R_i, we detail how to aggregate these *bba*'s in order to get the final belief concerning the reviewer classification. We combine these *bba*'s through the Dempster combination rule to obtain the whole *bba* that represent the evidence of the K-nearest Neighbors regarding the class of the reviewer. Hence, this global mass function m is calculated as such:

$$m_{R_i} = m_{R_i,R_1} \oplus m_{R_i,R_2} \oplus .... \oplus m_{R_i,R_K} \tag{18}$$

Decision making

In order to determine the membership of the reviewer R_i to one of the classes of Ω, we apply the pignistic probability $BetP$. Therefore, the classification decision is made either the reviewer is a spammer or innocent. For this, we select the class that has the grater value of $BetP$ as the final classification.

4 Experimentation and Results

The evaluation in the fake reviews detection problem was always a challenging issue due to the unavailability of the true real world growth data and variability of the features also the classification methods used by the different related works which can lead to unsafe comparison in this field.

Data description

In order to test our method performance, we use two datasets collected from yelp.com. These datasets represent the more complete, largest, the more diversified and general purpose labeled datasets that are available today for the spam review detection field. They are labeled through the classification based on the yelp filter which has been used in various previous works [2, 10, 16, 19, 26] as ground truth in favor of its efficient detection algorithm based on experts judgment and on various behavioral features. Table 1 introduces the datasets content where the percentages indicate the filtered fake reviews (not recommended) also the spammers reviewers.
The YelpNYC dataset contains reviews of restaurants located in New York City; the Zip dataset is bigger than the YelpNYC datasets, since it includes businesses in different regions of the U.S., such that New Jersey, New York, Vermont and Connecticut. The strong points of these datasets are:

- The high number of reviews per user, which facilities to modeling of the behavioral features of each reviewer.
- The divers kinds of entities reviewed, i.e., hotels and restaurants
- Above all, the datasets hold just basic information, such as the content, label, rating, and date of each review, connected to the reviewer who generated them. Thanks to the over-specific information, we can generalize the proposed method to different review sites.

Table 1. Datasets description

Datasets	Reviews (filtered %)	Reviewers (Spammer %)	Services (Restaurant or hotel)
YelpZip	608,598 (13.22%)	260,277 (23.91%)	5,044
YelpNYC	359,052 (10.27%)	160,225 (17.79%)	923

Evaluation Criteria

We rely on the three following criteria to evaluate our method performance: Accuracy, precision and recall, they can be defined as Eqs.19, 20, 21 respectively where TP, TN, FP, FN denote True Positive, True Negative, False Positive and False Negative respectively.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{19}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{20}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{21}$$

Experimental results

First of all, we apply the frequent itemset mining FIM, where I is the set of all reviewer ids in our two datasets. Each transaction is the set of the reviewer ids who have reviewed a particular hotel or restaurant. Thus, each hotel or restaurant generates a transaction of reviewer ids. By mining frequent itemsets, we find groups of reviewers who have reviewed multiple restaurants or hotels together. Then, we rely on the Maximal Frequent Itemset Mining (MFIM) to spot groups with maximal size in order to focus on the worst spamming activities. In the YelpZip dataset we found 74,364 candidate groups and 50,050 candidate groups for the YelpNYC dataset. We use $k = 3$ for our proposed approach.

Trying to ensure a safe comparison, we compare our method named Evidential Group Spammers Detection (EGSD) with two previous works in which authors rely on the FIM technique to generate the candidate groups and almost the same features used in our work. The first method introduced in [13] Detecting Group Review Spam (DGRS) used the FIM to generate candidate groups then computed the different indicators value and use the SVM rank algorithm to rank them, the other method proposed in [14] we focus on the Ranking Group Spam algorithm (GSRank) which rely on an iterative algorithm to effectively

Table 2. Comparative results

Evaluation Criteria	Accuracy			Precision			Recall		
Methods	DGRS	GSRank	EGSD	DGRS	GSRank	EGSD	DGRS	GSRank	EGSD
YelpZip	65%	78%	**85%**	70%	76%	**83.5%**	71%	74%	**86%**
YelpNYC	60%	74%	**84.3%**	62%	76.5%	**83.55%**	61.3%	77.2%	**85%**

rank the group spammers. The results are reported in the table 2.

Our method achieves the best performance detection according to accuracy, precision and recall over-passing the compared methods. We record at best an accuracy improvement over 10% in both YelpZip and YelpNYC data-sets compared to DGRS and over 7% compared to GSRank.

5 Conclusion

In this work, we tackle the group spammer review detection problem which become a real issue to the online rating systems and we propose a novel approach that aims to distinguish between the spammer and the innocent reviewers while taking into account the uncertainty in the different suspicious behavioral group spammer indicators. Experimental study on a real-world datasets against several state-of-the-art approaches verifies the effectiveness and efficiency of our method. Our proposed approach can be useful for different reviews sites in various fields. As future work, we aim to introduce other features to further improve the detection.

References

1. Akoglu, L., Chandy, R., Faloutsos, C.: Opinion fraud detection in online reviews by network effects. Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM, 13, 2-11 (2013)
2. Ben Khalifa, M., Elouedi, Z., Lefèvre, E. Fake reviews detection based on both the review and the reviewer features under belief function theory. The 16th international conference Applied Computing (AC'2019), 123-130 (2019)
3. Ben Khalifa, M., Elouedi, Z., Lefèvre, E.: Spammers detection based on reviewers' behaviors under belief function theory. The 32nd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE'2019). Springer International Publishing, 642-653 (2019)
4. Ben Khalifa, M., Elouedi, Z., Lefèvre, E.: Multiple criteria fake reviews detection using belief function theory. The 18th International Conference on intelligent systems design and applications (ISDA'2018). Springer International Publishing, 315-324 (2018)
5. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat.38, 325-339 (1967)
6. Denoeux, T.: A K-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Trans. Syst. Man Cybern. 25(5), 804–813 (1995)

7. Deng, X., Chen, R.: Sentiment analysis based online restaurants fake reviews hype detection. Web Technologies and Applications, 1-10 (2014)
8. Fayazbakhsh, S., Sinha, J.: Review spam detection: A network-based approach. Final Project Report: CSE 590 (Data Mining and Networks) (2012)
9. Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., Ghosh, R.: Exploiting burstiness in reviews for review spammer detection. Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM, 13, 175-184 (2013)
10. Fontanarava, J., Pasi, G., Viviani, M.: Feature Analysis for Fake Review Detection through Supervised Classification. Proceedings of the International Conference on Data Science and Advanced Analytics, 658-666 (2017).
11. Heydari, A., Tavakoli, M., Ismail, Z., Salim, N.: Leveraging quality metrics in voting model based thread retrieval. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 10 (1), 117-123 (2016)
12. Lim, P., Nguyen, V., Jindal, N., Liu, B., Lauw, H. : Detecting product review spammers using rating behaviors. Proceedings of the 19th ACM international conference on information and knowledge management, 939-948 (2010)
13. Mukherjee, A., Liu, B., Wang, J., Glance, N., Jindal, N. Detecting Group Review Spam. Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, ACM 978-1-4503-0637-9/11/03 (2011)
14. Mukherjee, A., Liu, B., Glance, N.: Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st international conference on world wide web, ACM, New York, 191–200 (2012)
15. Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M.: Spotting opinion spammers using behavioral footprints. Proceedings of the ACM international conference on knowledge discovery and data mining, 632-640 (2013)
16. Mukherjee, A., Venkataraman, V., Liu, B., Glance, N.: What Yelp Fake Review Filter Might Be Doing. Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM, 409-418 (2013)
17. Ong, T., Mannino, M., Gregg, D.: Linguistic characteristics of shill reviews. Electronic Commerce Research and Applications, 13 (2), 69-78 (2014)
18. Pan, L., Zhenning, X., Jun, A., Fei, W.: Identifying indicators of fake reviews based on spammer's behavior features. Proceedings of the IEEE International Conference on Software Quality, Reliability and Security Companion, QRS-C, 396-403 (2017)
19. Rayana, S., Akoglu, L.: Collective opinion spam detection: Bridging review networks and metadata. Proceedings of the 21th International Conference on Knowledge Discovery and Data Mining, ACM SIGKDD, 985-994 (2015)
20. Shafer, G.: A Mathematical Theory of Evidence, vol. 1. Princeton University Press (1976)
21. Smets, P.: The canonical decomposition of a weighted belief. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1896-1901 (1995)
22. Smets, P.: The transferable belief model for quantified belief representation. In: Smets, P. (ed.) Quantified Representation of Uncertainty and Imprecision, 267-301. Springer, Dordrecht (1998)
23. Wang, G., Xie, S., Liu, B., Yu, P. S.: Review graph based online store review spammer detection. Proceedings of 11th international conference on data mining, ICDM, 1242-1247 (2011)
24. Wang, Z., Hou, T., Song, D., Li, Z., Kong, T.: Detecting review spammer groups via bipartite graph projection. Comput J 59(6):861–874 (2016)

25. Wang, Z., Gu, S., Zhao, X., X, Xu.: Graph-based review spammer group detection. Knowledge and Information Systems, vol. 55, no. 3, pp. 571–597 (2017)
26. Wang, Z., Gu, S., Xu, X.: GSLDA: LDA-based group spamming detection in product reviews. Appl Intell 48, 3094–3107 (2018).
27. Xu, C., Zhang, J.: Towards collusive fraud detection in online reviews. In: 2015 IEEE international conference on data mining, ICDM Atlantic City, 1051–1056 (2015)
28. Xu, C., Zhang, J., Chang, K., Long. C.: Uncovering collusive spammers in chinese review websites. Proceedings of the 22nd ACM international conference on conference on information and knowledge management, ACM, New York, pp 979–988 (2013)