FAKE REVIEWS DETECTION BASED ON BOTH THE REVIEW AND THE REVIEWER FEATURES UNDER BELIEF FUNCTION THEORY

Malika Ben Khalifa, Zied Elouedi, Eric Lefevre

FAKE REVIEWS DETECTION BASED ON BOTH THE REVIEW AND THE REVIEWER FEATURES UNDER BELIEF FUNCTION THEORY

Malika Ben Khalifa[1;2], Zied Elouedi[1], and Eric Lefèvre[2]
1 Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisia
2 Univ. Artois, EA 3926, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Béthune, F-62400, France

ABSTRACT

The online reviews play an increasingly spreading role in consumer purchasing decisions and they are also considered as one of the most powerful source of information for companies. Due to this attraction, manufacturers and retailers rely on spammers to promote their own products and demote the competitors' one by posting fake reviews. Therefore, it is essential to detect deceptive reviews in order to ensure customers confidence and to maintain companies' fair competition. To tackle this problem, we propose a new approach able to spot spam reviews relying both on the rating reviews and the different spammers' indicators under the belief function framework. This method treats uncertainty in the given reviews also in the reviewers' information to take into account each reviewer spamicity when making decision. Experiments are conducted on two real-world review data-sets from Yelp.com with filtered (spam) and recommended (non-spam) reviews to demonstrate our method the effectiveness.

KEYWORDS

e-Commerce, Online reviews, Spammers, Fake reviews, Uncertainty, Belief function theory.

1. INTRODUCTION

In the recent years, the huge use of the internet changes the way people communicate and share their opinions. Online opinions are expressed as posts, comments, reviews or tweets in different websites. These online reviews have overturned the traditional consumer purchasing path. Checking them before making a purchase becomes a permanent habit. Hence, they represent the primary factor in a customer decision to purchase a product or service. However, reviews are more than just a way for customers to gather information, but also a powerful source information for companies. These latter have enlisted spammers to post positive deceptive reviews in order to elevate their products and subsequently write negative reviews to downgrade competitor of brand or company. Therefore, positive opinions bring significant financial gains for both business and individuals. However, negative reviews do not only cause financial loss, but also damage the companies e-reputation. Since, these fraudulent activities make the online reviews untruthful and unreliable, the fake reviews detection becomes more and more essential in order to maintain the readers confidence, to protect e-commerce and e-business and to ensure a fair competition between companies and bands. For these reasons, this challenging problem attracts significant researchers in the last years. They have developed several spam review detection methods in which the major task is distinguishing between trustful and deceptive reviews.

Several methods and approaches have been proposed to detect spam reviews. Most of them rely on the review content using the linguistic aspects and feeling. Moreover, several techniques used the individual words from the review text as features (Jindal and Liu, 2008; Mukherjee et al, 2013) while some others are based on the syntactic and lexical features. It is important to mention that most of the methods based only on the review content are not not sufficiently effective for review spam detection cause of the lack of any distinguishing words (features) that can give a definitive clue for classification of reviews as real or fake. Consequently, detecting spammers can improve detection of fake reviews, since many spammers share the same profile characteristics and activity patterns. For this reason, we notice the existing of various spammer detection

methods in which the graph-theory have been used to treat the relationships between the reviewer, their written reviews and the reviewed stores and most of them have shown promising results (Akoglu et al, 2013; Wang et al, 2011). Moreover, different features extracted from the reviewer characteristics and behavioral patterns have been used in several works (Fei et al, 2013; Lim et al, 2010; Mukherjee et al, 2013; Pan et al, 2017). Furthermore, combining spam review detection and spammer detection by analyzing their behaviors is more effective solution for detecting review spam than either approach alone. In this way, we cite the proposed methods in (Fontanarava et al, 2017; Rayana and Akoglu, 2015), that exploit both relational data and metadata of reviewers and reviews. Results prove that this kind of methods outperform all others.

Furthermore, all these previous methods display some weaknesses basically related to their inability to deal with the uncertainty in the different reviewers' information also in the given reviews which are often imperfect and imprecise. Disregarding such uncertainty may deeply sway the detection. Hence, handling the uncertainty when dealing with the fake reviewers' detection task becomes a prevalent interest. In this paper, we propose a new approach aiming to detect spam reviews by treating the uncertainty on both the rating reviews and the reviewers' spammer indicators under the belief function theory. This theory offers flexibility in representing and managing different types of imperfection. Accordingly, our method handles imperfections in different inputs both reviews and reviewers information in order to distinguish between fake and genuine reviews. The rest of this paper is structured as follows: In Section 2, we elucidate the fundamental concepts of the belief function theory. Then, Section 3 details our proposed method. Experimental study is discussed in Section 4. Finally, we conclude our work and we propose some future works in Section 5.

2. BELIEF FUNCTION THEORY

The belief function theory is one of the useful theories that handles uncertain knowledge. It was introduced by Shafer (1976) as a model to manage beliefs.

2.1 Basic concepts

The frame of discernment Ω is a finite and exhaustive set of different events associated to a given problem. 2^{Ω} is the power set of Ω that contains all possible hypotheses and it is defined by: $2^{\Omega} = \{A: A \subseteq \Omega\}$. A basic belief assignment (bba) or a belief mass is defined as a function from 2^{Ω} to [0,1] that represents the degree of belief given to an element A such that: $\sum_{A \subseteq \Omega} \boldsymbol{m^{\Omega}(A) = 1}$. A focal element A is a set of hypotheses with positive mass value $m^{\Omega}(A) > 0$. Various kinds of $bba's$ have been proposed by Smets (1995) in order to express special situations of uncertainty. Here, we underline 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 $\omega_i \in \Omega$.
- Simple support function: In this case, the bba focal elements are $\{A, \Omega\}$. A simple support function is defined as the following equation:

$$m^{\Omega}(X) = \begin{cases} \omega & if\ X = \Omega \\ 1 - \omega & if\ \ X = A\ for\ some\ A \subset \Omega \\ 0 & Otherwise \end{cases}$$

Where A is the focus and $\omega \in [0,1]$.

Moreover, the discounting operation (Ling and Rudd, 1989) allows us to update experts beliefs by taking into consideration their reliability through the degree of trust (1-α) given to each expert with $\alpha \in [0,1]$ is the discount rate. Accordingly, the discounted bba, becomes: $\begin{cases} m^{\Omega}(A) = (1-\alpha)\, m^{\Omega}(A) & \forall\, A \subset \Omega \\ m^{\Omega}(\Omega) = \alpha + (1-\alpha)\, m^{\Omega}(\Omega) \end{cases}$

Various numbers of combination rules have been proposed in the framework of belief function to aggregate a set of $bba's$ provided by pieces of evidence from different experts. Let $m_1^{\Omega}\ and\ m_2^{\Omega}$ 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 settled in (Smets, 1992), denoted by ⓒ and defined as:

$$m_1^{\Omega} ⓒ m_2^{\Omega}(A) = \sum_{B \cap C = A} m_1^{\Omega}(B)\, m_2^{\Omega}(C)$$

2. *Dempster's rule of combination:* This combination rule is a normalized version of the conjunctive rule (Dempster, 1967).

It is denoted by $\oplus$ and defined as:

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

3. *The combination with adapted conflict rule (CWAC):* This combination (Lefèvre and Elouedi, 2013) is an adaptive weighting between the two previous combination rules acting like the conjunctive rule if bbas are opposite and as the Dempster rule otherwise. They use the notion of dissimilarity that is obtained through a distance measure, to ensure this adaptation between all sources. The CWAC is formulated as follows:

$$m^{\Omega}_{\textcircled{\leftrightarrow}}(A) = (\textcircled{\leftrightarrow} m_i^{\Omega})(A) = D_{max} m^{\Omega}_{\textcircled{\cap}}(A) + (1 - D_{max}) m^{\Omega}_{\oplus}(A)$$

Where $\boldsymbol{D_{max}}$ is the maximal value of all the distances, it can be used to find out if at least one of the sources is opposite to the others and thus it may be defined by: $\boldsymbol{D_{max} = \max[d(m_i^{\Omega}, m_j^{\Omega})]}$, where $\boldsymbol{i \in [1, M]}$, $\boldsymbol{j \in [1, N]}$, M is the total number of mass functions and $\boldsymbol{d(m_i^{\Omega}, m_j^{\Omega})}$ is is the distance measure proposed by Jousselme et al. (2001):

$$\boldsymbol{d(m_1^{\Omega}, m_2^{\Omega}) = \sqrt{\frac{1}{2}(m_1^{\Omega} - m_2^{\Omega})^t D (m_1^{\Omega} - m_2^{\Omega})}}$$

Where D is the Jaccard index defined by: $\boldsymbol{D(E,F)} = \begin{cases} \boldsymbol{1} & \boldsymbol{if\ E = F = \emptyset} \\ \frac{|E \cap F|}{|E \cup F|} & \boldsymbol{\forall E, F \in 2^{\Omega} \setminus \emptyset} \end{cases}$

Frequently, we need to fuse two $bba's$ $m_1^{\Omega_1}$ and $m_2^{\Omega_2}$ that have not the same frame of discernment. So, we apply the vacuous extension of the belief function which extend the frames of discernment Ω_1 and Ω_2, corresponding to the mass functions $m_1^{\Omega_1}$ and $m_2^{\Omega_2}$, to the product space $\Omega = \Omega_1 \times \Omega_2$.
The vacuous extension operation, denoted by ↑ and defined such that:

$$m^{\Omega_1 \uparrow \Omega_1 \times \Omega_2}(B) = m^{\Omega_1}(A) \quad if \quad B = A \times \Omega_2$$

Where $A \subseteq \Omega_1$, $B \subseteq \Omega_1 \times \Omega_2$. It transforms each mass to the cylindrical extension to B to $\Omega_1 \times \Omega_2$.
Furthermore, the multi-valued mapping may be used to determine relation between two disjoint frames of discernment Ω_1 and Ω_2 (Dempster, 1967). This operation, denoted τ, allows us to join together two different frames of discernment the subsets $B \subseteq \Omega_2$ that can match through τ to be a subset $A \subseteq \Omega_1$:

$$m_{\tau}(A) = \sum_{\tau(B) = A} m(B)$$

The belief function framework offers various solutions to ensure the decision making. We present the pignistic probabilities, used in our work, denoted by *BetP* and defined as:

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

3. FAKE REVIEWS DETECTION BASED ON BOTH THE REVIEW AND THE REVIEWER FEATURES UNDER BELIEF FUNCTION THEORY

In this section, we present our novel proposed approach which tries to handle uncertainty in both the reviews and reviewers information in order to distinguish between honest and fake reviews. Our method is composed from three parts, in the first one we rely on the rating reviews by evaluating their trustfulness through their compatibility with all others which is inspired from (Ben Khalifa et al., 2018). In the second part and for the purpose of improving detection performance, we propose to take into consideration each reviewers' reliability by treating their behaviors (Ben Khalifa et al., 2019). Hence, we adopt the belief function theory to model

uncertainty in different imperfect review and reviewer information. Once the review and the reviewer trustfulness are represented by two mass functions, we combine them in the third part in order to make the more suitable decision. These three main parts of our approach are detailed in depth. In the following sections, we consider a dataset of N reviews illustrated by different rating vote values between 1 and 5 stars, each vote V_i is given by a reviewer denoted R_i where i is the id of the corresponding one.

3.1 Modeling the review trustworthiness

We propose to model imperfection in the rating reviews under the belief function framework by modeling each vote V_i through a mass function (i.e. bba) m_i^{Ω} with $\Omega = \{1,2,3,4,5\}$ where each element defines the rating reviews given by each reviewer.

3.1.1 Modeling the uncertain review

Generally and whatever the reviewer is spammer or innocent, we think that it is so difficult to provide an exact global vote to evaluate a product or a service. We think that the reviewer may provide an uncertain vote to a value close. Hence, we propose to model this uncertainty relative to the given vote by the vote, the vote +1 and the vote -1 represented by k and transformed into a mass function defined as: $m_{ik}^{\Omega}(\{k\}) = 1$ where $k \in \{V_i, V_{i+1}, V_{i-1}\}$. In the upper extreme case (i.e. $V_i = 5$), we model the vote and the vote-1 where $k \in \{V_i, V_{i-1}\}$ and in the lower one (i.e. $V_i = 1$) we model the vote and the vote+1 where $k \in \{V_i, V_{i+1}\}$.

Then, we propose to take into account the reliability degree of the vote V_i relying on his similarity to all others given votes. Thus, we model this reliability by $(1 - \alpha_i)$ where α_i is its discounting factor. We calculate α_i as follows: $\alpha_i = \frac{Number\ of\ votes\ different\ from\ the\ current\ vote\ of R_i}{Total\ votes'number}$

We apply the discounting operation to transform the vote into a simple support function.

Furthermore, we propose to take into account the gap between the given vote value V_i and its $(1 - \beta_{ik})$ corresponding represented values (vote, vote+1, vote-1) denoted by k, to not treat them in the same way, by where β_{ik} is its discounting factor and it is calculated as follows: $\beta_{ik} = \frac{|V_i - k|}{The\ maximum\ vote\ value}$

Therefore, each simple support function associated to each vote V_i is weakened by its relative reliability degree $(1 - \beta_{ik})$ using the discounting operation.

After that, we combine each three discounted $bbas$ (two in the extreme cases) modeling each vote using the Dempster rule in order to model each given uncertain vote by one global bba m_i^{Ω} which takes into account both the uncertainty and the reliability of this rating review.

3.1.2 Measuring the compatibility of the review with all the other ones

We propose to evaluate each rating review given by a distinct reviewer by comparing it with all the others reviewers votes'. Thus, for each vote, we aggregate all the others' using the CWAC combination rules which cope with the conflicts in these different bba and gives to as one combined bba which illustrates the whole reviews rating except the current one. Then, we measure the similarity between each rating review and all the others by calculating the distance $d(m_i^{\Omega}, m_{ci}^{\Omega})$ that separate them using the distance of Jousselme.

3.1.3 Modeling the review into trustful or not trustful

The calculated distance defines the compatibility degree between the vote and all the other ones' and it implies that more the distance value decreases more the vote is trustful. Therefore, we propose to transform each distance into a new bba with $\Theta = \{t, \bar{t}\}$ (t for trustful and $\bar{t}$ for not trustful) as the following equation:

$$\begin{cases} m^{\Theta}(\{t\}) = \gamma * (1 - \frac{1}{1 + e^{-a.ds + \frac{a}{2}}}) \\ m^{\Theta}(\{\bar{t}\}) = \gamma * \frac{1}{1 + e^{-a.ds + \frac{a}{2}}} \\ m^{\Theta}(\Theta) = 1 - \gamma \end{cases}$$

where $ds = d(m_i^{\Omega}, m_{ci}^{\Omega})$, $a = 10$ and $\gamma = \frac{The\ standard\ deviation\ of\ all\ votes}{The\ maximum\ value\ of\ the\ standard\ deviation}$

Through these steps, we model each review trustworthiness by the mass function m_i^{Θ} under the frame of discernment $\Theta = \{t, \bar{t}\}$.

3.2 Modeling the reviewer spamicity

The review compatibility with all other reviews is one of the most used indicators to spot deceptive reviews. However, spammers may post a huge number of fake reviews in order to overturn the majority of the given reviews. Therefore, it is crucial to rely also on the reviewers' spamicity to improve the distinguishing between the fake and honest reviews. In this way, we propose to model uncertainty in the different spammer behaviors extracted from the reviewers profiles information. We represent each reviewer R_i by two mass functions namely; the reviewer reputation $m_{RR_i}^{\Omega_s}$ and the second one is to model the reviewer helpfulness $m_{RH_i}^{\Omega_s}$ with $\Omega_s = \{S, \bar{S}\}$ where S is spammer and $\bar{S}$ is not spammer.

3.2.1 Modeling the reviewer reputation

In the spammer review detection field, it has been proved that spammers are excepted to post a huge number of reviews to limited intended products or services in short time span, say in two or three days, in order to over-qualify or damage some specific products. However, the genuine reviewers post their opinion when they have actually bought new products or used new services. Hence, their reviews are always steadily over time interval and depend on the number of used products or services. So, we can construct the reviewer reputation through these two spammers' indicators. Therefore, we propose to verify the reviewing history associated to each reviewer $Hist_{R_i}$ defined as the set of all past reviews given by the reviewer R_i for n discrete products. Each reviewer average proliferation is calculated through the sum of his given reviews and divided by the total number of reviewed products n through the following equation: $AvgP(R_i) = \frac{Hist_{R_i}}{n}$

We assume that if $AvgP(R_i) > 3$ the reviewer is considered as a potential spammer since generally ordinary reviewers do not give more than three reviews per product. Thus, the reviewer reputation is then modeled by a certain bba as follows: $m_{RR_i}^{\Omega_s}(\{S\}) = 1$ eLse $m_{RR_i}^{\Omega_s}(\{\bar{S}\}) = 1$

Moreover, we propose to check if the reviews are given in a short time of interval or are scattered during the reviewing history. Therefore, we consider the most used time interval; three days and we measure the brust spamicity degree α_i by calculating the sum of the reviews' number given in less than three days divided by the total number of reviews by each reviewer denoted by TNR_i as follows:

$$\alpha_i = \frac{Number\ of\ reviews\ given\ by\ R_i\ in\ less\ than\ 3\ days}{TNR_i}$$

In order to take into account the brust spamicity degree, we weaken each reviewer reputation mass function by its corresponding reliability degree (i.e., $(1 - \alpha_i)$ or α_i) using the discounting operation.

As a result, we obtain the discounted bba ${}^{\alpha} m_{RH_i}^{\Omega_s}$ which presents the reviewer reputation relying on two important spammer indicators namely; the reviewer's average proliferation and the brust spamicity.

3.2.2 Modeling the reviewer helpfulness

The reviewer helpfulness is also one of the important indicators to detect spammers. Accordingly, we propose to use the Number of Helpful Reviews (NHR) associated to each reviewer to verify if the reviewer post helpful reviews or unhelpful one to mislead readers.

Therefore, if $(NHR_i = 0)$, the reviewer is suspicious to be spammer, thus we model the reviewer helpfulness by a certain bba as follows: $m_{RH_i}^{\Omega_s}(\{S\}) = 1$ else $m_{RH_i}^{\Omega_s}(\{\bar{S}\}) = 1$

We propose to measure the non-helpfulness degree β_i of each reviewer R_i in order to not consider all the reviewers who give helpful reviews in the same way. So, we penalize each reviewer helpfulness mass by its discounting factor β_i calculated as follows: $\beta_i = \frac{TNR_i - NHR_i}{TNR_i}$

Then, we apply the discounting operation in order to transform the bba into a simple support function ${}^{\beta} m_{RH_i}^{\Omega_s}$ Thus, we take into consideration the helpfulness degree.

Commonly, the honest reviewers are completely satisfied or dissatisfied by the used products or services. Therefore, they will not usually post extreme rating. The spammers continuously rely on the extreme ratings (Mukherjee et al, 2013), either highest (5 stars) or lowest (1 star), for the purpose of reaching their objectives of speedily increasing or bringing down, respectively, the mean score of a product. So, the reviewer may have a lot of helpful reviews but if they are full of extreme rating, his chances of being genuine reviewer will absolutely decrease. In order to take this fact into consideration, we calculate the extreme rating degree

denoted γ_i, corresponding to each reviewer R_i, which is considered as the discounting factor calculated by the number of the extreme rating divided by the total number of reviews given by each reviewer TNR_i as the following equation: $\gamma_i = \frac{NER_i}{TNR_i}$, where NER_i is the extreme reviews' number (i.e., $NER_i \in \{1,5\}$) given by each reviewer R_i. Then, each simple support function represented the reviewer helpfulness ${}^{\beta}m_{RH_i}^{\Omega_s}$ is weakened another time by its relative reliability degree (i.e., (1-γ_i) or γ_i) through the discounting operation.

Thus, the resulting discounted ${}^{\beta\gamma}m_{RH_i}^{\Omega_s}$ modeled the reviewer helpfulness based on both the reviewer helpfulness degree and extreme rating.

3.2.3 Combining the both reviewer reputation and helpfulness

In the interest of illustrating the whole spamicity for each reviewer, we aggregate the reviewer $bba's$ reputation ${}^{\alpha}m_{RH_i}^{\Omega_s}$ with ${}^{\beta\gamma}m_{RH_i}^{\Omega_s}$ his helpfulness using the Dempster rule of combination under the frame of discernmentΩ_s. The joint mass function $m_{RT_i}^{\Omega_s}$ illustrates each reviewer's the spamicity degree.

3.2 Distinguishing between the fake and the genuine reviews

As highlighter before, combining spam review and spammer review detection by analyzing their behaviors become the most effective solution to detect fake reviews.

Therefore, we propose to combine both the review trustworthiness modeled by m_i^{Θ} with the reviewer spamicity represented by $m_{RT_i}^{\Omega_s}$ in order to make a suitable decision. For this, we have to apply the following steps:

3.2.1 Modeling both the reviewer and the review trustworthiness

In order to express the review and the reviewer information through one bba, we apply the following steps:

- Define Ω_{RR} as the global frame of discernment relative to the review trustworthiness and the reviewer spamicity. It defines the cross product of the two different frame Θ and Ω_s denoted by: $\Omega_{RR} = \Theta \times \Omega_s$
- Extend all the review trustworthiness and the reviewer spamicity $bbas$, respectively m_i^{Θ} and $m_{RS_i}^{\Omega_s}$ to the global frame of discernment Ω_{RR} to get new $bbas$ $m_i^{\Theta\uparrow\Omega_{RR}}$ and $m_i^{\Omega_s\uparrow\Omega_{RR}}$ using the vacuous extension.
- Combine the different extend $bbas$ using the Dempster rule of combination. $m_i^{\Omega_{RR}} = m_i^{\Theta\uparrow\Omega_{RR}} \oplus m_i^{\Omega_s\uparrow\Omega_{RR}}$

Finally $m_i^{\Omega_{RR}}$ represents both the review and the reviewer trustworthiness.

3.2.2 Review and reviewer trustworthiness transfer

The next step is to transfer the combined $m_i^{\Omega_{RR}}$ under the product space Ω_{RR} to the frame of discernment $\Theta_D = \{f, \bar{f}\}$ in order to make decision by modeling the reviews into fake or not fake.

In spam reviews detection field, all the reviews given by the spammers are considered as fake reviews, even if they provide a review compatible with all the given ones, it is considered as fake reviews because spammers are not real consumers and they usually use this method to avoid being detected by the spammer detection methods. For that, a multi-valued operation, denoted τ is applied. The function $\tau:\ \Omega_{RR}$ to 2^{Θ_D} rounds up event pairs as follows:

- Masses of event couples with at least an element $\{S\}$ spammer are transferred to fake $f \subseteq \Theta_D$ as:
  - $m_\tau(\{f\}) = \sum_{\tau(SR_i)=f} m_i^{\Omega_{RR}}(SR_i),\ (SR_i = A \times S) \subseteq \Omega_{RR}$
- Masses of event couples with at least an element $\{\bar{S}\}$ not spammer are transferred to fake $f \subseteq \Theta_D$ as:
  - $m_\tau(\{\bar{f}\}) = \sum_{\tau(SR_i)=\bar{f}} m_i^{\Omega_{RR}}(SR_i),\ (SR_i = A \times \bar{S}) \subseteq \Omega_{RR}$
- Masses of event couples with at least no element $\{S,\ \bar{S}\}$ are transferred to Θ_D as:
  - $m_\tau(\Theta_D) = \sum_{\tau(SR_i)=\Theta_D} m_i^{\Omega_{RR}}(SR_i),\ (SR_i = A \times \Omega_s) \subseteq \Omega_{RR}$

3.2.3 Decision making

Now that we transferred all $bbas$ modeling both the reviews and the reviewer information to the decision fame of discernment Θ_D in order to differentiate between the fake and the genuine reviews. Thus, we apply the pignistic probability $BetP$. Finally, the $BetP$ with the grater value will be considered as the final decision.

4. EXPERIMENTATION AND RESULTS

4.1 Evaluation protocol

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

4.1.1 Datasets description

In this study, we use two datasets collected and used in (Mukherjee et al, 2013; Rayana and Akoglu, 2015) from yelp.com. These datasets with near-ground-truth are considered as the largest, complete, full of information about the reviews, reviewers and also the reviewed services. They are labeled through the yelp filter which has been used in many previous works (Fontanarava et al, 2017; Mukherjee et al, 2013; Rayana and Akoglu, 2015; Heydari et al, 2016) as a ground truth thanks to its efficient detection method based on various behavioral features, where recommended (Not filtered) reviews correspond to genuine reviews, and not recommended (filtered) reviews correspond to fake ones. Table 1 presents the datasets statistics in which we indicate also the percentages of filtered (Fake) reviews. Due to the huge number of reviews, we random sample the two datasets with 10% from the total number of reviews and we evaluate our method through the three following criteria: Accuracy, precision and recall.

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

4.1.3 Experimental results

As our method proposes a classifier which can distinguish between fake and genuine reviews under an uncertain context. We propose to compare it with the state-of-art baselines classifiers; the Support Vector Machine (SVM) and the Naive Bayes (NB) used by most of the spam detection methods (Lim et al, 2010; Fei et al, 2013; Mukherjee et al, 2013; Rayana and Akoglu, 2015). Moreover, we compare our method with the proposed uncertain classifier Belief Fake Reviews Detection (BFRD) in (Ben Khalifa et al, 2018) which relies only on the review rating information. The results are reported in the table 2.

Table 2. Comparative results

Evaluation Criteria	Accuracy				Precision				Recall			
Methods	NB	SVM	BFRD	Our method	NB	SVM	BFRD	Our method	NB	SVM	BFRD	Our method
YelpZip	55%	61%	72%	**87%**	57%	64%	76%	**89%**	56%	62%	74%	**88%**
YelpNYC	60%	66%	78%	**91.5%**	62%	68%	80%	**92.55%**	61.3%	66.8%	79.2%	**90%**

Our method achieves the best performance detection according to accuracy, precision and recall over-passing the baseline classifier. We record at best an accuracy improvement over 25% in both yelpZip and yelpNYC data-sets compared to NB and over 20% compared to SVM. Moreover, the improvement records be-tween the two uncertain classifier (over 13%) shows the importance of combining both the review and the

reviewer features in this field. Despite the fact that our approach is based on fewer indicators than yelp's filter method, we obtain competitive results (over 90%) thanks to our method ability in handling uncertainty within the different inputs. These encouraging results push us to integrate more behavioral features in our future work that we could improve our results and obtain identical or even better performance than yelp filter.

5. CONCLUSION

In this study, we tackle the spam review detection problem and proposed a novel method that deals with the uncertainty in both the review and the reviewer centric features and relying on the spammer behavior indicators to spot fake reviews. Experimentations are conducted on two large real labeled datasets from yelp.com, the promising results obtained show the ability of our method in distinguishing between the fake reviews and the genuine ones.

REFERENCES

Akoglu, L. et al, 2013. Opinion fraud detection in online reviews by network effects. *Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM*, 13, pp. 2-11.

Ben Khalifa, M. et al, 2018. Fake reviews detection under belief function framework. *Proceedings of the International Conference on Advanced Intelligent System and Informatics (AISI)*, pp. 395-404.

Ben Khalifa, M. et al, 2018. Spammers detection based on reviewers' behaviors under belief function theory. *Proceedings of the International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems.* (To appear)

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.*38, 325-339

Lefèvre, E. and Elouedi, Z., 2013. How to preserve the confict as an alarm in the combination of belief functions? *Decis. Support Syst.*56, pp. 326-333.

Fei, G. et al, 2013. Exploiting burstiness in reviews for review spammer detection. *Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM*, 13, pp. 175-184.

Fontanarava, J. et al, 2017. Feature Analysis for Fake Review Detection through Supervised Classification. *Proceedings of the International Conference on Data Science and Advanced Analytics*. pp. 658-666.

Heydari, A. et al, 2016. 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), pp.117-123.

Jindal, N. and Liu, B., 2008. Opinion spam and analysis. *Proceedings of the 2008 International Conference on Web Search and Data Mining. ACM*, pp. 219-230.

Jousselme, A.-L. et al, 2001. A new distance between two bodies of evidence. *Inf. Fusion* 2(2), pp. 91-101.

Lim, P. et al, 2010. Detecting product review spammers using rating behaviors. *Proceedings of the 19th ACM international conference on information and knowledge management*. pp.939-948.

Ling, X. And Rudd, W., 1989. Combining opinions from several experts. *Applied Artificial Intelligence an International Journal*, 3 (4), pp. 439-452.

Mukherjee, A. et al, 2013. Spotting opinion spammers using behavioral footprints. *Proceedings of the ACM interna- tional conference on knowledge discovery and data mining*. pp. 632-640.

Mukherjee, A. et al, 2013. What Yelp Fake Review Filter Might Be Doing. *Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM*. pp. 409-418.

Pan, L. et al, 2017. Identifying indicators of fake reviews based on spammers behavior features. *Proceedings of the IEEE International Conference on Software Quality, Reliability and Security Companion, QRS-C*. pp. 396-403

Rayana, S. and Akoglu, L., 2015. Collective opinion spam detection: Bridging review networks and metadata. *Proceedings of the 21th International Conference on Knowl- edge Discovery and Data Mining, ACM SIGKDD*. pp. 985-994.

Shafer, G., 1976. *A Mathematical Theory of Evidence*, vol. 1. Princeton University Press.

Smets, P., 1995. The canonical decomposition of a weighted belief. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. pp. 1896-1901.

Smets, P., 1998. The transferable belief model for quantified belief representation. *Smets, P. (ed.) Quantified Representation of Uncertainty and Imprecision*. pp. 267-301.

Wang, G. et al, 2011. Review graph based online store review spammer detection. *Proceedings of 11th international conference on data mining, ICDM*. pp. 1242-1247.