Multiple criteria fake reviews detection using belief function theory

Malika Ben Khalifa[1], 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. Checking online reviews before making a purchase becomes a permanent habit. Hence, online consumer reviews, product and services play an increasingly spreading role in consumer purchasing decisions. Unfortunately, the importance of advertising and the attraction of profit have led to the appearance of fake reviews in order to mislead readers. Considering that the reviews are generally imperfect, the spam reviews detection becomes one of the most important problems. To tackle this problem, we propose a new method of multi-criteria fake reviews under belief function theory. This approach treats the uncertainty in the rating reviewers' given to multiple evaluation criteria, takes into account the similarity between all provided reviews and deals with missing data. We evaluate our method through artificial datasets. Then, we use a real dataset to validate it. The results prove that the proposed approach is a useful solution for the fake reviews detection problem.

Keywords: Online reviews, Multi-criteria evaluation, Fake reviews, Uncertainty, Belief function theory.

1 Introduction

The Web 2.0 movement has increased the use of the online reviews which are posted on social media, opinion-sharing websites, blogs, forms and merchant websites. User generated reviews in which they express their opinions, attitudes, and feeling regards to their experiences with some products or services.
Such opinions straightforwardly influence future customers purchase decisions and consequently the success of businesses. Companies and products with high rating reviews attracted more customers and can yield monetary gains. However negative reviews damage the companies' reputation and cause financial loss.
Unfortunately, driven by the desire for profit, spammers have tried to cheat the online review system by creating deceptive reviews in order to mislead readers. The fake review is an unreal review because it does not reflect the opinion's real consumer. It can be positive to over qualify some products and to promote some services and brands, or negative to distort the perceived quality of the competitors' products and damage the companies' reputation.

Therefore, the fake reviews detection becomes an essential task in order to stop this online threat, to save companies reputation, to maintain a trustful environment between consumer and producer and to protect online reviews system. As a consequence, this non-trivial and challenging problem attracted significant researchers' attention. They proposed several detecting online fraud techniques which serve to spot the fake reviews from the genuine ones. These methods use various heuristics, such as duplicated reviews [9] or fake reviews detection from non-experts [12] to generate a reference data. A large number of works tries to detect fake reviews relying on review text information as the style of writing [2, 7]. Other methods have focused on the detection of group spammers cause of their power in the deviation of the products or services identification through their members whose can publish numerous fake reviews in a few minutes. A set of eight indicators has elucidated in [11] that tries to catch the behavior of the group members such as time and rating deviation. Researchers establish some techniques to detect spammers. Most of these are networks reviews based approaches [1, 6] with tree types of nodes namely: review, reviewer and stores. Although the fake reviews detection is an uncertain problem, no one of these previous works is able to manage uncertainty in the reviews. However, we have proposed a preliminary related work dealing with the uncertainty but in only one overall criterion [3]. We think that the reviewer can better express his opinion through different evaluation criteria. Hence in this paper, we propose a new method, Multi-Criteria Belief Fake Reviews Detection (MC-BFRD), to handle uncertainty in the reviews' ratings given by each reviewer to several evaluation criteria using the belief function theory. Indeed, this theory is able to handle uncertainty and manage imperfections. It can manipulate various pieces of information from different reviewers and also allows to deal with partial and total ignorance. The rest of this paper is organized as follows: In Section 2, we present the fundamental concepts of the belief function theory. Then, in Section 3, we elucidated our proposed method (MC-BFRD). Experimentation evaluations are discussed in Section 4. Finally, we conclude in Section 5.

2 Belief Function Theory

The belief function theory was introduced by Shafer [13] as a model to represent beliefs. It is considered as an efficient tool able to deal with uncertainty and to manage several types of imperfection. Various models have been proposed from this theory. One of the most used is the Transferable Belief Model (TBM) [14] that we adopt in our method.

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 defined as a function from 2^Ω to $[0, 1]$ that represents the degree of belief given to an element A such that: $\sum_{A \subseteq \Omega} m^{\Omega}(A) = 1$. A focal element A is a set of hypotheses with positive mass value $m^{\Omega}(A) > 0$.

2.2 Discounting operation

Discounting [13] allows us to weaken the masses by the *discount rate* $\alpha \in [0, 1]$ such that $(1 - \alpha)$ is the degree of confidence of the source.
The *bba* is discounted as follows:

$$\begin{cases} {}^{\alpha}m^{\Omega}(A) = (1-\alpha)m(A) & \forall A \subset \Omega, \\ {}^{\alpha}m^{\Omega}(\Omega) = \alpha + (1-\alpha)m(\Omega). \end{cases} \tag{1}$$

2.3 Combination Rules

Two *bbas*, m_1^{Ω} and m_2^{Ω} induced by two distinct and cognitively independent reliable sources of information can be combined by several rules of combination. Each rule has its specificities and its characteristics. Then, we will introduce some of the most used ones.

1. *Conjunctive rule* (CRC)**:** It is introduced in [15], denoted by $\bigcirc\!\!\!\!\cap$ and defined as:
$$m_1^{\Omega} \bigcirc\!\!\!\!\cap\, m_2^{\Omega}(A) = \sum_{B \cap C = A} m_1^{\Omega}(B) m_2^{\Omega}(C) \tag{2}$$
2. *Dempster's rule of combination***:** It is the normalized version of the conjunctive rule that denies the mass on the empty set [4]. It is denoted by $\oplus$ and defined as:
$$m_1^{\Omega} \oplus m_2^{\Omega}(A) = \begin{cases} \frac{m_1^{\Omega} \bigcirc\!\!\!\!\cap\, m_2^{\Omega}(A)}{1 - m_1^{\Omega} \bigcirc\!\!\!\!\cap\, m_2^{\Omega}(\emptyset)} & \text{if} \quad A \neq \emptyset, \forall A \subseteq \Omega, \\ 0 & otherwise. \end{cases} \tag{3}$$
3. *The combination with adapted conflict rule* $(CWAC)$**:** This combination [5] 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:
$$m_{\bigcirc\!\!\!\!\leftrightarrow}^{\Omega}(A) = (\bigcirc\!\!\!\!\leftrightarrow m_i^{\Omega})(A) = D_{max} m_{\bigcirc\!\!\!\!\cap}^{\Omega}(A) + (1 - D_{max}) m_{\oplus}^{\Omega}(A) \tag{4}$$
where D_{max} represents 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:
$$D_{max} = max[d(m_i^{\Omega}, m_j^{\Omega})], \tag{5}$$
where $i \in [1, M]$, $j \in [1, M]$, M is the total number of mass functions and $d(m_i^{\Omega}, m_j^{\Omega})$ is the distance measure proposed by Jousselme [10]:
$$d(m_1^{\Omega}, m_2^{\Omega}) = \sqrt{\frac{1}{2}(m_1^{\Omega} - m_2^{\Omega})^t D (m_1^{\Omega} - m_2^{\Omega})}, \tag{6}$$

where D is the Jaccard index defined by:

$$D(E,F) = \begin{cases} 1 & \text{if} \quad E = F = \emptyset, \\ \frac{|E \cap F|}{|E \cup F|} & \forall E, F \in 2^{\Omega} \backslash \emptyset \end{cases} \tag{7}$$

2.4 Vacuous extension

In some cases we need to combine two *bbas* $m_1^{\Omega_1}$ and $m_2^{\Omega_2}$ that have not the same frame of discernment. So, we use 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 a joint compatible frame of discernment Ω defined as follows:

$$\Omega = \Omega_1 \times \Omega_2 \tag{8}$$

The extended mass function of any evidential value of the extended focal element B, denoted by ${}^{\Omega_1 \uparrow \Omega_1 \times \Omega_2}$, is defined as follows:

$$m_1^{\Omega_1 \uparrow \Omega_1 \times \Omega_2}(B) = m_1^{\Omega}(A) \quad if \quad B = A \times \Omega_2 \tag{9}$$

Where $A \subseteq \Omega_1$, $B \subseteq \Omega_1 \times \Omega_2$.
It transforms each mass to the cylindrical extension B to $\Omega_1 \times \Omega_2$.

2.5 Decision Making

The Transferable Belief Model (TBM) proposed by [15] is composed by two main levels: The credal level where evidence is defined by *bbas* and combined and the pignistic level where *bbas* are represented by probability functions called the pignistic probabilities denoted by $BetP$ and defined as:

$$BetP(B) = \sum_{A \subseteq \Omega} \frac{|A \cap B|}{|A|} \frac{m^{\Omega}(A)}{(1 - m^{\Omega}(\emptyset))} \quad \forall\, B \in \Omega \tag{10}$$

3 Multi-Criteria Belief Fake Reviews Detection (MC-BFRD)

Given a dataset of N reviewers and q evaluation criteria. We consider that each reviewer R_i judges a product or a service by giving votes V_{ij}, with values between 1 and 5 stars (respectively poor, below average, average, good and excellent), to each proposed criterion C_j.

3.1 General case

In the general case of our MC-BFRD, we assume that each reviewer assigns votes to all the evaluation criteria without missing anyone. Our method follows four main steps will be established and will be explained in-depth.

Step 1: Modeling reviewer's opinion by mass functions
As we adopt the belief function theory to model uncertainty in votes, each vote V_{ij} corresponding to each criterion C_j, with j in $[1, q]$ the number of evaluation criterion, will be transformed into a mass function (i.e. bba) $m_{ik}^{\Omega_j}$ with $\Omega_j = \{1, 2, 3, 4, 5\}$ where one of the elements in Ω_j represents the stars' number given by each reviewer R_i.
We propose to model this uncertainty by considering the vote, the vote-1 and the vote+1 for each given rating criterion review. In the upper extreme case, we only consider the vote and the vote-1 and in the lower one we model the vote and the vote+1. We transform each vote value V_{ij} given by the reviewer R_i corresponding to each criterion C_j into *bbas* defined as follows:

$m_{ik}^{\Omega_j}(\{k\}) = 1$ where $k \in \{V_i, V_{i+1}, V_{i-1}\}$

- In the upper extreme case, $k \in \{V_i, V_{i-1}\}$.
- In the lower extreme case, $k \in \{V_i, V_{i+1}\}$.

We propose to model the reliability degree of the reviewer R_i on each criterion C_j by $(1 - \alpha_{ij})$ where α_{ij} is its discounting factor. Its value is between $[0, 1]$, if $\alpha_{ij} = 0$ the reviewer is totally reliable and if $\alpha_{ij} = 1$, it means that the reviewer is totally unreliable and it will not be taken into consideration. α_{ij} is calculated as follows:

$$\alpha_{ij} = \frac{\text{Number of votes different from the current vote of } R_i \text{ on } C_j}{\text{Total votes' number for each criterion } C_j} \tag{11}$$

So, each vote V_{ij} is weakened by its relative reliability degree $(1 - \alpha_{ij})$ using the discounting operation (Eq.1). Thus, the reliability of the reviewer for each criterion will be taken into consideration.
In addition, we propose to take into account the distance between i^{th} value of the given vote for the corresponding criterion C_j denoted by V_{ij} and the values that model it (vote, vote+1 and vote-1) noted k and represented by $m_{ik}^{\Omega_j}$. This distance is modeled by β_{ijk} which will be considered as a weakening factor. Its value is between $[0, 1]$, if $\beta_{ijk} = 0$ then the model vote corresponds to the one provided and if $\beta_{ijk} = 1$ it means that the vote model is very different from the one given. The discounting factor β_{ijk} is calculated as follows:

$$\beta_{ijk} = \frac{|V_{ij} - k|}{\text{The maximum vote value}} \tag{12}$$

As a result, each simple support function associated to each vote given to each criterion is weakened by its relative reliability degree $(1 - \beta_{ijk})$ using the discounting operation (Eq.1).
Then, we aggregate the discounted *bbas* representing the provided vote using the Dempster rule (Eq.3) in order to represent each vote given to each criterion through one combined *bba*, $m_i^{\Omega_j}$ with $i = 1, \ldots, N$ and $j = 1, \ldots, q$.
Moreover, in order to express the reviewer's opinion through one *bba*, we have to apply the following steps:

– We define Ω_c as the global frame of discernment relative to all criteria. It represents the cross product of the different Ω_j denoted by:

$$\Omega_c = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_q \tag{13}$$

– We extend the different $bba's$ $m_i^{\Omega_j}$ to the global frame of all criteria Ω_c to get different $bba's$ $m_i^{\Omega_j \uparrow \Omega_c}$
– We combine different extend $bba's$ using the Dempster rule of combination.

$$m_i^{\Omega_c} = m_i^{\Omega_1 \uparrow \Omega_c} \oplus m_i^{\Omega_2 \uparrow \Omega_c} \oplus \ldots \oplus m_i^{\Omega_q \uparrow \Omega_c} \tag{14}$$

Finally, $m_i^{\Omega_c}$ represents the reviewer's opinion given through the different votes criteria.

Step 2: Distance between the current reviewer's opinion and all the other opinions' aggregation

To evaluate the opinions provided by each reviewer R_i through the different given criteria C_j, we compared it to all other reviewers' opinions as follows:
For each vote given to each criterion, we aggregate all the other reviewers' votes given to the same criterion represented by *bbas* using the CWAC combination rule (Eq.4) chosen, because it can manage the conflict in different votes. The result of this combination is one *bba* $m_{ic}^{\Omega_j}$, which represents the whole reviewers' votes for each criterion except the current one as follows:
$m_{ic}^{\Omega_j} = m_1^{\Omega_j} \circledlr m_2^{\Omega_j} \circledlr \ldots \circledlr m_{i-1}^{\Omega_j} \circledlr m_{i+1}^{\Omega_j} \circledlr \ldots \circledlr m_N^{\Omega_j}$.
As a result, for each reviewer we have q $bba's$ $m_{ic}^{\Omega_j}$, we propose to aggregate them in order to represent all reviewers' opinions except the current one in one joint *bba*. So, we must firstly extend them to the global frame of criteria Ω_c to get different $bba's$ $m_{ic}^{\Omega_j \uparrow \Omega_c}$.
Then, we combine different extend $bba's$ using the Dempster rule of combination:

$$m_{ic}^{\Omega_c} = m_i^{\Omega_1 \uparrow \Omega_c} \oplus m_i^{\Omega_2 \uparrow \Omega_c} \oplus \ldots \oplus m_i^{\Omega_q \uparrow \Omega_c} \tag{15}$$

This *bba* represents the all reviewers' votes except the current one given to all the q different evaluation criteria. Then, for each reviewer we calculate the distance $d(m_i^{\Omega_c}, m_{ic}^{\Omega_c})$ using the distance of Jousselme (Eq.6), in order to measure the similarity between each review's opinion and all others on all the different evaluation criteria.

Step 3: Construction of a new *bba* modeling the reviewer's opinion into fake or not fake

The distance founded in the previous step represents the degree of compatibility between each reviewer opinion and all the other ones'. The closer they are, more the reviewer is reliable. So, we propose to transform the distance value into a new *bba* with $\Theta = \{f, \bar{f}\}$ f= fake and $\bar{f}$= not fake as the following equation:

$$\begin{cases} m_i^{\Theta}(\{f\}) = \gamma * \frac{1}{1+e^{-a.ds+\frac{a}{2}}} \\ m_i^{\Theta}(\{\bar{f}\}) = \gamma * (1 - \frac{1}{1+e^{-a.ds+\frac{a}{2}}}) \\ m_i^{\Theta}(\Theta) = 1 - \gamma \end{cases} \quad (16)$$

with $ds = d(m_i^{\Omega}, m_{ic}^{\Omega})$, $\gamma = \frac{\text{The standard deviation of all votes}}{\text{The maximum value of the standard deviation}}$ and $a = 10$ where a corresponds to the slope at the point of inflection. Taking the equation $-ax + a/2$ allows the inflection point to be $x = 0.5$.

Step 4: Decision Making
The decision process is assured by the pignistic probability $BetP$. We select the hypothesis with the greater value of $BetP$ and we considered it as the final decision.

3.2 Missing values' case

In our MC-BFRD general case, we assume that reviewers give votes to all the evaluation criteria, but we can frequently find reviewers who vote some evaluation criteria and miss others. So, we propose to adapt our method in order to deal with the incomplete data. Therefore, we update the first step, which consists in modeling reviewer's opinion by mass functions, as follows:
In the case of incomplete data, we add the case of the vacuous vote, represented by a vacuous *bba* $m_i^{\Omega}(\Omega) = 1$.
Furthermore, we continue with the same instructions of the first step.
Moreover, the next steps will be exactly identical as those of the general case, namely distance between the current reviewer's opinion and all the other opinions' aggregation, construction of a new *bba* modeling the reviewer's opinion into fake or not fake and decision making.

4 MC-BFRD Experimentation

In the fake reviews detection, the evaluation is one of the most challenging problems considering the unavailability of the labeled dataset because it is not obvious to distinguish between the fake and the real reviews. Thus, researchers have used human evaluation in most previous work. However, human evaluation remains subjective since different evaluators often have different levels of tolerance. In this paper, we conducted experiments on four different artificial datasets then we propose to test our method behavior through one real dataset.

4.1 MC-BFRD evaluation

We propose to evaluate our method through the creation of four different artificial dataset that we labeled artificially by fake or not fake. Our labeling is based on the majority of votes' combinations given to the various proposed criteria.

Experimental protocol

Given four datasets in which reviewers judge an hotel by giving votes to three different criteria in order to express their opinion on four different hotels.

- The first dataset composed by 220 reviewers that votes hotel with 3 different criteria. The majority of the opinions are composed by 3, 4 and 5 stars that we labeled as not fake and the others as fake.
- The second dataset contains 420 reviewers that express their opinion through three different criteria. We label the majority reviewers opinions' composed by 3, 4 and 5 stars as not fake and the rest as fake.
- The third dataset composed by 270 reviewers that give votes to some criteria to evaluate one hotel. The majority of the opinions are composed by 1 and 5 stars that we labeled as not fake and 2, 3 and 4 stars as fake.
- The fourth dataset contains 265 reviewers that judge an hotel by voting three different criteria. This dataset provides incomplete data. We label the majority of reviewers opinions composed by 1 and 2 stars as not fake and the others as the fake ones.

Table 1 presents the description of the different datasets.

Table 1. Dataset description

Dataset	Number of reviewers	Labeled as fake	Labeled as not fake
Dataset1	220 reviewers	20 reviewers' opinions	200 reviewers' opinions
Dataset2	420 reviewers	70 reviewers' opinions	350 reviewers' opinions
Dataset3	270 reviewers	20 reviewers' opinions	250 reviewers' opinions
Dataset4	265 reviewers	25 reviewers' opinions	240 reviewers' opinions

The most used classification evaluation measures is accuracy. It is defined as follows:

$$Accuracy = \frac{\text{The number of well classified instances}}{\text{The total number of classified instances}} \tag{17}$$

Experimental results

Table 2 presents for the average accuracies of the four executed datasets with the majority voting (MV) and our MC-BFRD. It demonstrates that our method gives a higher accuracy in all tested datasets that reach 92% with the third dataset while the majority vote method provides a modest accuracy where the best one was 57%. It proves that our method gives better results in comparison of those obtained with the majority vote with a large gap.

Table 2. Average accuracies (MV vs. MC-BFRD)

Dataset	MV accuracy	MC-BFRD accuracy
Dataset1	0.44	0.90
Dataset2	0.50	0.80
Dataset3	0.57	0.92
Dataset4	0.40	0.79

4.2 MC-BFRD behavior validation

In order to evaluate the performance of our MC-BFRD, we conduct a real database extracted from Tripadvisor. It consists of a 1550 reviewers who express their opinion on "Melia Caribe Tropical hotel" by writing reviews, giving an overall and vote other criteria like rooms, location, cleanliness, business and services. Moreover, this dataset contains some missing values. In our MC-BFRD, we deal with the different evaluation criteria. Figure 1 details the obtained results with the Tripadvisor dataset in which our proposed method uncovers 703 fake reviews and 847 genuine ones.

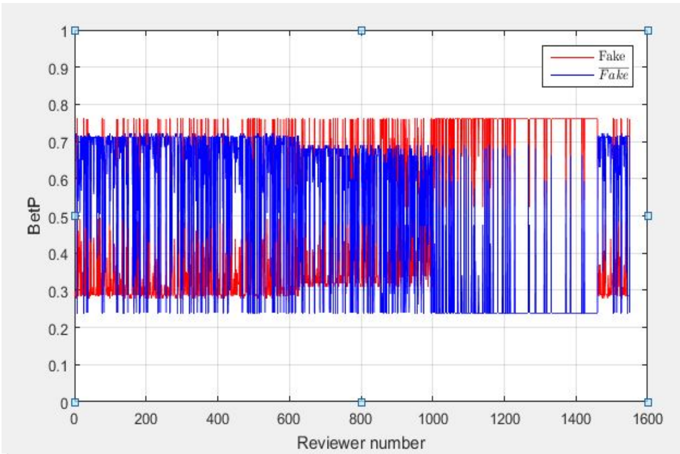

Fig. 1. *BetP* results of Tripadvisor dataset

5 Conclusion

In this paper, we tackle the fake reviews detection problem in an uncertain context using the belief function framework. We proposed to deal with uncertainty in the different given rating reviews to multiple evaluation criteria and and takes into consideration the compatibility with all other ones. In fact, our new method shows effectiveness in uncovering the fake reviews from the honest ones. Moreover, this approach can be applied in several fields such as *e*-commerce and *e*-business and help different websites to detect fraudulent rating reviews. As future work, we will integrate some other notions like reviewers' trustiness and we will also take into consideration the different spam indicators.

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. Banerjee, S., Chua, A. Y. K.: Applauses in hotel reviews: Genuine or deceptive? Proceedings of science and information conference (SAI), 938-942 (2014)
3. Ben Khalifa, M., Elouedi, Z., Lefèvre, E.: Fake reviews detection under belief function framework. Proceedings of the International Conference on Advanced Intelligent System and Informatics (AISI), 395-404 (2018)
4. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat.38, 325-339 (1967)
5. Lefèvre, E., Elouedi, Z.: How to preserve the confict as an alarm in the combination of belief functions? Decis. Support Syst.56, 326-333 (2013)
6. Fayazbakhsh, S., Sinha, J.: Review spam detection: A network-based approach. Final Project Report: CSE 590 (Data Mining and Networks) (2012)
7. Fusilier, D. H., Montes-y-Gómez, M. M., Rosso, P., Cabrera, R. G.: Detection of opinion spam with character n-grams. Computational linguistics and intelligent text processing, 285-294 (2015)
8. 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)
9. Jindal, N., Liu, B.: Opinion spam and analysis. Proceedings of international conference on web search and data mining, 219-230 (2008).
10. Jousselme, A.-L., Grenier, D., Bossé, É.: A new distance between two bodies of evidence. Inf. Fusion 2(2), 91-101 (2001)
11. 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)
12. Ott, M., Yejin, C., Claire, C., Jerey, T. H.: Finding deceptive opinion spam by any stretch of the imagination. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 1, 309-319 (2011)
13. Shafer, G.: A Mathematical Theory of Evidence, vol. 1. Princeton University Press (1976)
14. Smets, P.: The combination of evidence in the transferable belief model. IEEE Trans. Pattern Anal. Mach. Intell. 12(5), 447-458 (1990)
15. Smets, P.: The transferable belief model for quantified belief representation. In: Smets, P. (ed.) Quantified Representation of Uncertainty and Imprecision, pp. 267-301. Springer, Dordrecht (1998)