A Machine Learning approach for the Identification of the Deceptive Reviews in the Hospitality Sector using Unique Attributes and Sentiment Orientation

M.R. Martinez-Torresa [a,*] , S.L. Toral[b]

[a] Facultad de Ciencias Económicas y Empresariales. University of Seville (Spain), Av. de Ramón y Cajal, 1, 41018, Sevilla, Spain

[b] E. S. Ingenieros, University of Seville (Spain), Avda. Camino de Los Descubrimientos s/n, 41092, Seville, Spain

* Corresponding author. E-mail addresses: rmtorres@us.es (M.R. Martinez-Torres), storal@us.es (S.L. Toral)

Abstract

The popularity of online reviews is causing a huge impact on consumers' purchase intentions for goods and services. However, and hidden by the anonymity of the Internet, fraudsters can try to manipulate other consumers by posting fake reviews. Maintaining trust in online reviews require the development of automatic tools using machine learning approaches because of the huge volume of online opinions generated every day. This paper is focused on the hospitality sector and follows a content analysis approach based on a set of unique attributes and the sentiment orientation of reviews. The main contributions of the paper are i) a set of polarity-oriented unique attributes able to distinguish positive and negative deceptive and non-deceptive reviews and ii) the main topics associated to positive and negative deceptive and non-deceptive reviews. Findings reveal that positive and negative unique attributes lead to non-biased classifiers and that experience based reviews tend to be non-deceptive.

Keywords: Deceptive reviews; online reviews; unique attributes; sentiment orientation; classifiers

1. Introduction

Today, most of independent travel related booking is done online, and after consumption, travellers have the option of providing feedback in the form of online reviews. They are fast, up-to-date and publicly available, and they constitute the electronic version of traditional word-of-mouth (eWOM, electronic word-of-mouth) (Schuckert et al., 2014). In the case of the hospitality and tourism industry, consumers trust reviews as they are independent from official or corporate information (Zhu & Zhang, 2010), and they show the previous experience of other travellers using their own words (Toral et al., 2018). They assist the decision-making process of potential customers and, indirectly, encourage hospitality managers to improve their product or service quality.

However, all these benefits can be compromised by the increasing presence of deceptive reviews. In contrast to the voluntary and honest feedback provided by real consumers, fraudsters posting deceptive reviews pursuit the manipulation of other customers to artificially promote or devalue products and services. Recent studies estimate that around 25-30% of online reviews are deceptive reviews (Roberts, 2013; Li et al., 2014). Although there is a consensus in the fact that online reviews will continue growing in the future, the presence of fake reviews is a threat that can undermine consumer confidence on shared opinions (Chen et al., 2017).

The success of online reviews is based on their credibility, so they influence the attitude towards the product and the purchase intention whenever they come from a credible source (Shan, 2016; Banarjee et al., 2017). However, the concept of a credible source in Social Media is different to that of traditional word-of-mouth, where the source is someone belonging to the inner circle of the consumer. In the case of online reviews, users are anonymous and only identified by an alias. One possible source of credibility is the reviewer's reputation, which is rated by other users (Cheung et al., 2009). Typically, eWOM websites not only display the content and the author of the review, but also some other information like the system-generated profiles of reviewers (Martínez-Torres et al., 2018). The system-generated profile includes a brief information about the author posting the review as well as the community-rated reputation of reviewers, indicating the perceived usefulness of previously posted reviews and other products purchased or rated (Wu, 2013). Many eWOM websites such as Amazon, TripAdvisor or Ciao, allow users to vote on the "helpfulness" of posted reviews (Arenas-Marquez et al., 2014; Gonzalez-Rodriguez et al., 2016; Geetha et al., 2017). However, malicious users can easily manipulate online reputation systems, as the reputation is based on simple rules that can be distorted through false accounts that perform false scoring to artificially improve their trustworthiness (Kirilenko et al., 2019).

While there exist researches that study the reviewers' reputation to differentiate between deceptive and non-deceptive reviews, this paper follows a different approach that consists of analysing the content of reviews. The main challenge is that deceptive reviews always try to resemble honest reviews, so the aim of the paper is to identify those differences (features) that make possible a successful classification of deceptive and non-deceptive reviews, following a text-based machine learning approach. The body of reviews exhibit important differences depending on the sentiment polarity (positive or negative). Generally, negative deceptive opinions are more difficult to be detected than positive spam (Fusilier et al., 2015). Therefore, the sentiment polarity will also be considered when collecting differentiating features (Zhang et al., 2018). The hospitality sector will be analysed as a case of study. The advantage of following a text-based machine learning approach is twofold. First, it provides an automatic tool able to

process a huge volume of reviews. Second, it learns from a specific context. As a difference from other approaches, machine learning techniques learns the specific vocabulary of a specific sector or industry, so it can learn better than other approaches based on generic features.

The remainder of the paper is structured as follows: section 2 details the related work regarding the classification schemes for deceptive reviews detection. Section 3 presents the research question and hypotheses. Section 4 describes the case study based on a public dataset and the methodology followed for the identification of deceptive reviews and topics. Section 5 shows the obtained results and section 6 discusses how the research question and the proposed hypotheses are answered through these results. This section also includes the implications, limitations and future work. Finally, section 7 concludes the paper.

2. Related work

Previous literature has considered three main different approaches for obtaining relevant features of fake reviews: linguistic or review centric methods, reviewer centric, and network approaches (Crawford et al., 2015; Jiang et al., 2016; Bi et al, 2019).

The aim of *review centric approaches* is finding predictive deception cues in the content of a message. The simplest method of representing texts is the "bag of words" approach, where individual or small groups of words (n-grams) from the text combined with their TF (Term Frequency) or TF-IDF (Term Frequency-Inverse Document Frequency) values are used as features (Larcker & Zakolyukina 2012). Part of Speech (POS) tagging (Markowitz & Hancock, 2014), affective dimensions or location-based words (Hancock et al, 2013) can also provide frequency sets able to reveal linguistic cues of deception. Stylometric features are also used to identify fake reviewers based on the writing style traces embedded in their online comments (Shojaee et al., 2013) These lead to deep syntax analysis methods that focus on distinguishing rule categories (lexicalized or unlexicalized) for deception detection (Feng et al., 2012). Finally, semantic analysis can be used to find signals of truthfulness (Lau et al., 2011). The intuition is that a deceptive writer with no experience with a product or service (e.g., never visited the hotel in question) may fall into contradictions or omission of facts that are present in profiles on similar topics (Conroy et al., 2015).

Reviewer centric approaches focus on features collected from reviewer profile characteristics and behavioural patterns. Features such as the number of reviews, the percentage of positive reviews, the deviation from the average review rating, the review length, and the presence of similar reviews for different products by the same reviewer, or the variety of products or services where the reviewer is posting reviews, are considered (Mayzlin, 2014). All this information is part of the system-generated profile and can be easily collected in online reviews, as it is publicly available (Olmedilla et al., 2016a).

Finally, the *network approaches* refer to the analysis of interdependencies through the links or edges between objects (either reviewers or reviews) to obtain the behaviour of users in online reviews and eWOM websites. The interactions are modelled as a social network, and the micro and macro analysis can then reveal suspicious behaviours that can be associated to fraudsters. Ku et al. (2012) studied the role of the users' trust network from a micro perspective by considering the users' trust network as a 2-hop network. They define the trust intensity given by the size of the trust network as well as the average intensity of the 2-hop neighbours, which is the trust intensity of the members of the trust network. Both of them are positively related to

the reputation of the member. From a macro perspective, PageRank-like approaches solve suspicious node detection problem in large graphs from the ranking perspective, such as MailRank for spam ranking (Chirita et al., 2005) or FraudEagle for fraud ranking (Akoglu et al., 2013). In addition, from a macro perspective, density-based detection methods in graphs look for areas of higher density than the remainder of the graphs/data (Akoglu et al., 2010). Hybrid methods combining previous approaches have also been treated in the previous works. For instance, Barbado et al. (2019) added social features (friends, followers, votes…) to the reviewer centric features in the case of consumer electronics.

In this paper we follow a review centric approach. We are interested in obtaining the specific attributes of deceptive and non-deceptive reviews in the case of the hospitality sector, and considering the sentiment polarity of reviews.

Regarding the detection methods, supervised methods are clearly the most frequent methods reported in the literature: linear/logistic regression models, naive Bayesian models, SVM, nearest-neighbour algorithms (such as k-NN), least squares, ensembles of classifiers and multi-layer perceptrons (Zhang et al., 2012; Larcker & Zakolyukina, 2012; Jiang et al., 2016; Ahsan et al., 2016). There are many machine learning algorithms, so it is not easy the decision about which one is best. Additionally, each machine-learning algorithm has two types of model parameters: ordinary parameters, that are automatically optimized or learned in a model-training phase, and hyper-parameters, that are typically set by the user of a machine learning software tool manually before a machine-learning model is trained (Luo, 2016).

3. Research framework

Review centric approaches are based on a set of features given by a bag of words. Therefore, the performance of classifiers for the identification of deceptive reviews relies on the selection of relevant terms for this task. Some previous papers addressed the increased difficulty of identifying negative deceptive reviews (Ott et al., 2013; Fusilier et al., 2015; Chevalier & Mayzlin, 2006), which lead to biased classifiers. This result can be explained because negative deceptive reviews are more similar to truthful reviews, as they are mainly related to complains. The selection of features is based either on the relative frequency of terms (Teso et al., 2018) or on the experience of researchers (Do et al., 2006). As the identification of positive deceptive reviews is easier, classifiers get biased towards better classifying this set of reviews. There have been some attempts to improve feature selection. For instance, Agnihotri et al. (2017) propose a variable global feature selection scheme for automatic classification of text documents considering a minimum number in each class. The main problem of previous methods is the interpretation of results. The selection of features attending only to the frequency or normalized frequency of terms does not say anything about their discriminative properties among classes. Obviously, some of the terms discriminate the target classes as the performance of the classifier is good enough, but there are also many other terms with a very small contribution to the performance of the classifier. In a pure prediction problem, the presence of a high number of features is not a problem. But when interpreting the results in terms of the selected features, those terms with a small contribution make difficult to differentiate between the topics of deceptive and non-deceptive reviews. Gao et al. (2019) considered the characteristics of the review content and the reviewers' behaviour to identify deceptive reviews. They integrated sentiment analysis and the characteristics of reviewers, and utilized a feature-weighted model to describe the emotional intensity of the reviews and the importance of the characteristics of

the reviewers. Once they identified deceptive reviewers through their unreliability scores, they featured their reviews (deceptive reviews) as having a high emotional intensity value and being the text very similar to other reviews. Moreover, they identified that non-deceptive opinions use positive words to express their sentiments, while deceptive favorable opinions use more compliments, with a stronger sentiment opinion and more exaggerated languages.

In this paper we propose the use of the so-called unique features (Toral et al., 2018), which stands for those attributes that are uniquely associated to a given class among all the possible classes. The main advantage of using unique attributes is that they are the terms that account a major contribution to the classifier performance, that is, they are the terms with better discriminative properties. As the polarity of reviews has been demonstrated to lower the identification scoring of negative deceptive reviews, we will perform the unique feature selection considering positive and negative deceptive/non-deceptive reviews. Hence, we posit:

> *H1: The polarity-oriented selection of unique attributes keeps the performance of the classification of deceptive and non-deceptive reviews lowering the number of attributes.*
>
> *H2: The polarity-oriented selection of unique attributes improves the association with positive and negative deceptive and non-deceptive reviews*

Review centric approaches are specifically focused on the content of shared reviews. Works based on writing style consider POS and psycholinguistic features (Tausczik & Pennebaker, 2010). For instance, it has been argued that genuine reviews appeared less hyperbolic compared with deceptive ones (Banerjee & Chua, 2014). However, previous literature has demonstrated that, considering that a review is fake because it conveys an extreme opinion, is false (Li and Hitt 2008; Dellarocas and Wood 2008). In many cases, extreme opinions are posted as a result of a very good or bad experience. In such cases, honest reviewers can be very positive or very negative, the same than presumed deceptive reviews. Therefore, it is necessary to go deeper inside the content of reviews. According to Barsky and Labagh (1992), the most valued attributes by guests when visiting a hotel are connected to the “reception”, “employee attitudes”, “facilities”, “services” and “location”. Regarding complaints, they can be categorized into four groups: (1) “physical environment”, which refers to noise, décor, parking, view, atmosphere, ambience, accommodations, room location (Li et al., 2017), (2) “physical goods”, including food and beverage quality, climate control, temperature of the pool, elevator service, cleanliness, furniture condition, pool (Xu & Li, 2016; Hu et al., 2019), (3) “service & personnel”, which refers to reservation handling, management attitude, service speed, employee attitude, level of service (Lee & Hu, 2005), and (4) “expectations”, which includes elation to advertising, available facilities, package plan delivery, price-value (Berezina et al., 2016). Depending on the opinion’s sentiment, Ott et al. (2013) concluded that positive and negative deceptive reviews include less spatial details due to the ignorance of not having been there (in a hotel). The application of topics analysis to online reviews shows that sentiment orientation influences the number of a variety of topics. Mankad et al. (2016) found that negative reviews tend to focus on a small number of topics, whereas positive reviews tend to touch on a greater number of topics. Additionally, Hernández-Castañeda et al (2018) found that relevant topics for negative non-deceptive opinions include words that demerit things, while the relevant topics of positive non-deceptive opinions consist of words such as warmly, experience, greatest, restaurant, experience, central).

We suggest that some topics can be used to distinguish deceptive and non-deceptive reviews, but they are different depending on the sentiment polarity. More specifically, we hypothesize:

> *H3: Deceptive positive reviews emphasize hotel ubication while non-deceptive positive reviews highlight city characteristics*
>
> *H4: Deceptive positive reviews emphasize hotel and room characteristics while non-deceptive positive reviews are more focused on feelings and experiences* related to the stay at the hotel.
>
> *H5: Deceptive negative reviews emphasize physical and tangible inconveniences while non deceptive negative reviews are more focused on expectations and feelings.*

4. Case study and Methodology

The dataset comprises 800 honest reviews, and 800 deceptive reviews uniformly distributed across 20 popular hotels in Chicago (Ott et al., 2011; Ott et al., 2013). This dataset was selected for two reasons. First, reviews are annotated as deceptive, non-deceptive, positive and negative (400 reviews for each case). Annotated data is required for building classifiers able to distinguish between deceptive and non-deceptive reviews, as it is a supervised machine learning technique. Second, it is a public dataset widely cited by the scholarly community (Sun et al., 2013; Ong et al., 2014; Parapar et al., 2014).

The proposed methodology is depicted in Figure 1. Automatic text-based approach relies on the selection of a representative bag of words, but prior to the selection, it requires first a pre-processing stage consisting of cleaning the input data, which are the body of reviews written by reviewers. The pre-processing stage involves stop-words and punctuation removal, lower case conversion and stemming. Stop-words are words that do not carry information, such as prepositions, pronouns or articles. Eliminating stop-words helps to improve text processing performance (Yee Liau & Pei Tan, 2014). The aim of the subsequent stage is to account each word belonging to the bag of words, every time they appear within the body of the reviews. By removing punctuations and conducting a lower-case conversion, the text is homogenized. Finally, stemming is a process of transforming words into their roots. By removing derivational affixes, all the possible variants of a given word are accounted as the same word. In this paper, we use the traditional Porter Stemmer, which is the standard stemmer used in NLP and Information Retrieval tasks (Porter, 1980).

Following the scheme of Figure 1, the annotated dataset is randomly split into a training dataset (80%) and a test dataset (20 %). The train dataset is used to train a set of classifiers and the test dataset to report the accuracy of the trained classifiers. Classifiers are trained using as input values the TF-IDF of a bag of words. The TF-IDF is a normalized value that balance the frequency of words with its rarity along the corpus of documents (Youn Kim and Yoon 2013), so it is appropriate for discriminating classes of documents. Three options will be considered to collect the bag of words. The first one consists of selecting all the existing words. The second one consists of selecting the so-called unique attributes, which are those attributes that can be uniquely associated to one of the classes (Toral et al., 2018). Basically, the unique attribute identification consists of applying an ANOVA to the TF-IDF values followed by a Turkey's Honestly Significant Difference (HSD) test. In the context of this study, we have two classes, deceptive and non-deceptive reviews, so unique attributes are uniquely associated to one of

them. As a third option, we consider a polarity oriented unique attribute selection, which consists of extended the prior two classes to four classes, by adding the sentiment polarity. Thus, unique attributes are uniquely associated to positive deceptive reviews, positive non-deceptive reviews, negative deceptive reviews or negative non-deceptive reviews.

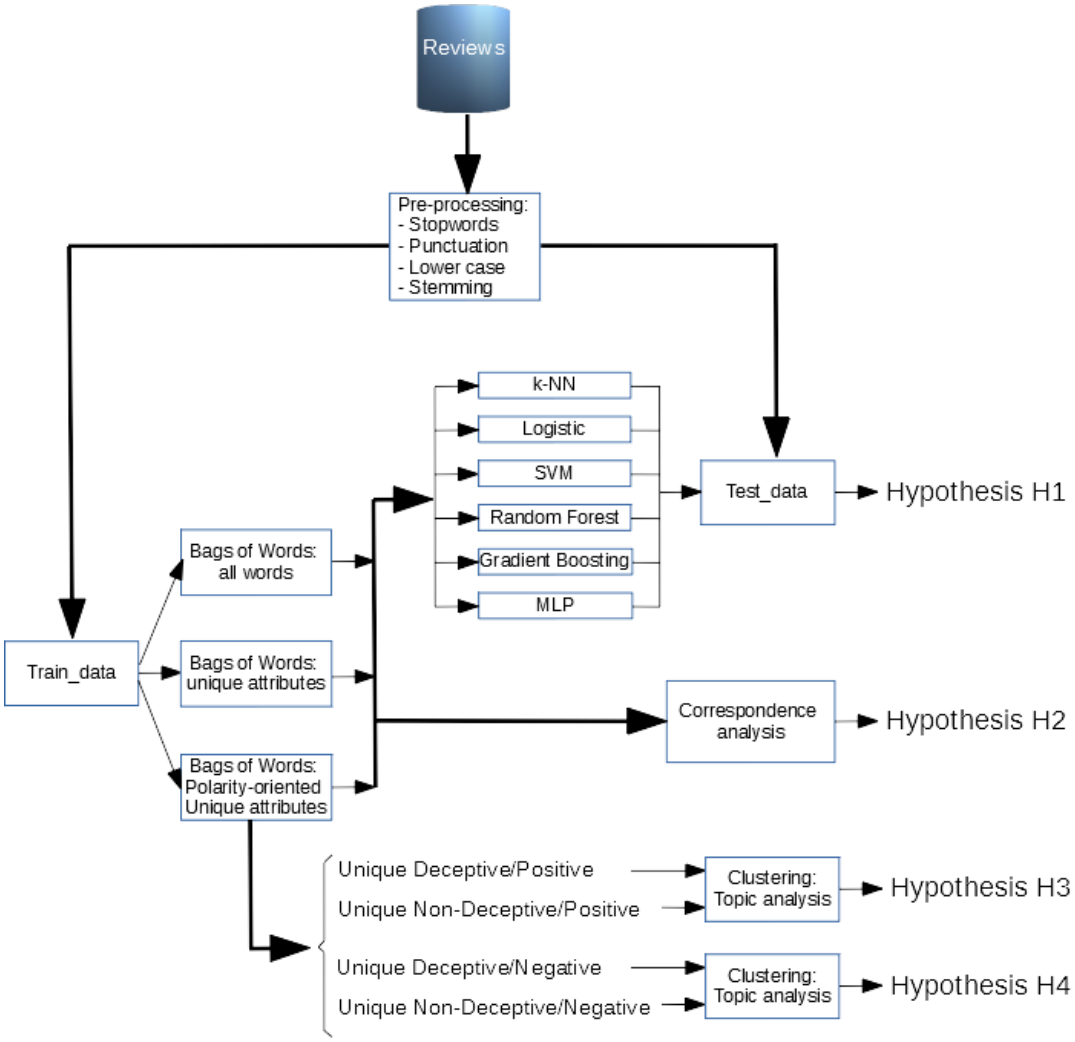


Figure 1. Block diagram of the proposed methodology.

Six different classifiers will be trained and tested using each bag of words. Precision and recall will be provided as the output metrics for classifiers. Both values are calculated using the confusion matrix illustrated in Figure 2 with the formulas of Eq. (1).

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

Figure 2. Confusion matrix

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \qquad Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (1)$$

Intuitively, the precision is the ability of the classifier not to label as positive a sample that is negative, while the recall is intuitively the ability of the classifier to find all the positive samples. The advantage of using precision and recall instead of simple accuracy is that they consider misclassified elements, so with both values it is possible to check if the classifier is biased toward one of the classes. A combination of precision and recall is given by F1-score, defined as:

$$F1\text{-}score = 2\ x\ \frac{Precision\ x\ Recall}{Precision+Recall} \quad (2)$$

The F1-score summarizes in a single value precision and recall, so it can be used for comparison purposes.

The proposed polarity-oriented of unique attributes provide four set of words with good discriminant properties among classes. They can be visualized using a correspondence analysis, which is a grouping method used for understanding similarities and association between variables and it has become popular for dimensional reduction and perceptual mapping (Whitlark and Smith 2001). As a second part of the methodology (bottom part of Figure 1), a clustering analysis will be applied to each one of the four set of words to obtain the main topics within each group. As a result, the topics for deceptive and non-deceptive positive reviews can be compared separately to the topics of deceptive and non-deceptive negative reviews.

5. Results

The 1600 annotated reviews were pre-processed following the steps previously detailed in the methodology section. Table 1 shows two examples of the resulting text after removing stop words and punctuations and after applying lower case conversion and stemming. The stemmed text preserves the root of words. In the second row of Table 1 it can be observed that the words 'stayed' and 'staying' of the original text column are stemmed as 'stay'. The same can be said for the second example for the words 'price' and 'priced' that are stemmed to the common root 'price'.

Table 1. Results of text pre-processing including stop words and punctuation removal, lower case conversion and stemming.

Original text	Stemmed text (6, 1588)
We stayed in the Conrad for 4 nights just before Thanksgiving. We had a corner room overlooking N Michigan Av and the Tribune Building. To say this is a 5 star hotel is to damn this place with faint praise - it is wonderful. The staff were unbelievably helpful. The standard of the room was superb - HD plasma screens, luxury bedlinens, iPod radio, huge bathroom. Location is unbeatable - right in the heart of everything - watched the whole Light Festival parade from the window. Breakfasts were excellent - no help yourself buffet here - with full and attentive table service. Would have no hesitation recommending or staying again in this hotel - 5 out of 5.	stay conrad night corner room overlook michigan tribun build say star hotel place faint wonder staff help standard room superb hd plasma luxuri ipod radio huge bathroom locat unbeat right heart everyth watch whole light window breakfast excel help buffet full attent tabl servic would hesitation recommend stay hotel
My stay at The Palmer House Hilton was less than satisfactory. I would have expected much for from such a	stay palmer hous hilton less satisfactori would expect much

'high end' hotel. The rooms were way over priced, and the service was less than satisfactory. At one point, I left my dirty towels on the floor to be picked up, and when room service came through, they were still there! On the other side of things, the cup I had left on my desk (which I was planning to use later) was gone. My coffee and my towels were also never replaced. I also was astounded that internet service did not come with the price of the room. Rather, it was a 'nominal fee'. After paying so much for a room, I had hoped to at least have complimentary wifi. My stay at the Palmer House Hilton left much to be wished for. I will not be staying there again.	high end hotel room way price servic less satisfactori point left dirti towel floor pick room servic came still side thing cup left desk plan use later gone coffe towel never replac internet servic come price room rather fee pay much room hope least complimentari wifi stay palmer hous hilton left much wish stay

The first bag of words consists of collecting all the stemmed words after the pre-processing stage but considering a minimum count of 20. That is, all those words that don't occur at least 20 times within the whole set of documents are discarded, as they are not good candidates as features for the classifiers due to its low frequency. The resulting number of words is 918.

The second bag of words consist of considering the unique attributes for the two classes, deceptive and non-deceptive reviews. The total number of unique attributes is 296, being 198 deceptive class attributes, and 98 non-deceptive class attributes.

The third bag of words corresponds to the polarity oriented unique attributes for the four classes, deceptive positive, non-deceptive positive, deceptive negative, and non-deceptive negative. The number of unique attributes is 134, with 29 for deceptive positive class, 24 for non-deceptive positive, 44 for deceptive negative, and 37 for non-deceptive negative.

For each of the three previous options, the TF-IDF value of words were calculated and provided as input to the classification stage. Six different classifiers were trained: k-NN, logistic, Support Vector Machines (SVM), Random Forest, Gradient Boosting, and Multi-Perceptron (MLP). The reason for choosing six different classifiers is because of the well-known No-Free-Lunch Theorem, a principle formulated by Wolpert and Macready (1997), that basically says that all machine learning or optimisation algorithms perform equally well when their performance is averaged against all possible datasets and objective functions. Therefore, there is no guarantee that one algorithm will perform better than others. Table 2 details the description, parameters and selected hyperparameters for the six proposed classifiers.

Table 2. Description, parameters and hyperparameters of classifiers.

Classifier	**Description: Parameters**	**Hyperparameters**
k-NN	k-nearest neighbour: non-parametric approach	k=3
Logistic	Logistic classifier: weights of features	-
SVM	Support Vector Machines: the support vectors, the Lagrange multiplier for each support vector	Regularization factor C=1; kernel=linear
RF	Random Forest: the input variable used at each internal node of a decision tree, the threshold value chosen at each internal node of a decision tree	The number of decision trees=500, the number of features to consider when looking for the best

		split=0,5 (maximum), maximum depth=10
GB	Gradient Boosting: the input variable used at each internal node of a decision tree, the threshold value chosen at each internal node of a decision tree	The number of decision trees=500, the minimum number of samples required to split an internal node=5, maximum depth=10
MLP	Multi-layer Perceptron: the weight on each edge and bias values	Number of layers=3, layer sizes=[16,8,8]

The 80% of the original dataset (randomly selected) was used to train the classifiers, and the other 20% was used to report the output metrics. Table 3 compares the performance of the classifiers for the three options of bag or words. The metrics of this table were computed using the test dataset. In addition to precision, recall and F1-score, this table also includes the number of False Positives (FP) and False negatives (FN), which correspond to misclassified elements. It can be noticed that the best result for the F1-score is provided by the Support Vector Machine (SVM) classifier (0.881, 0.873 and 0.833 for the three options of bag of words). This result proves that the performance of the classifier is almost the same while lowering the number of attributes from 918 in the case of all words to 296 in the case of unique attributes and to 134 in the case of polarity-oriented unique attributes. Therefore, and as posited by hypothesis H1, the polarity-oriented unique attributes keep the performance of classifiers when they discriminate between deceptive and non-deceptive reviews.

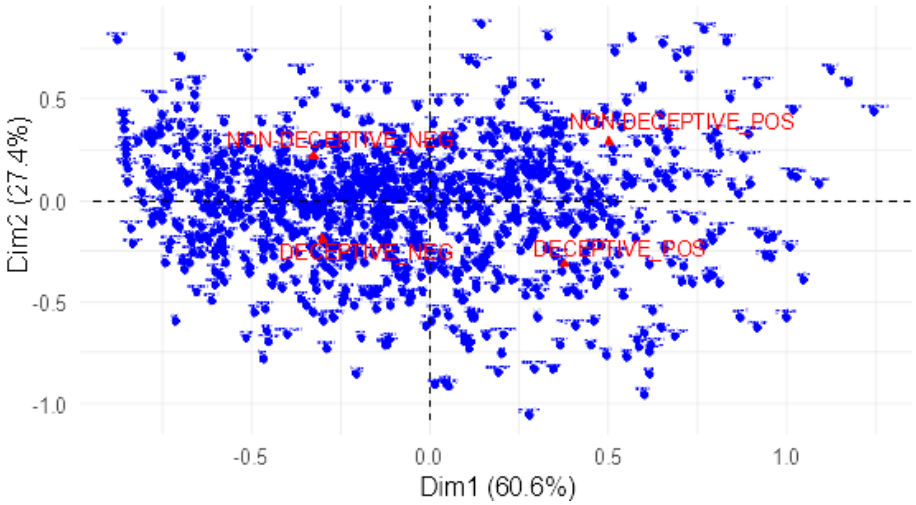
Table 3. Output metrics of selected classifiers for the three options of bag of words.

Bag of words	Classifier	FP	FN	Precision	Recall	F1-score
All words	k-NN	29	55	0.786	0.660	0.718
	Logistic	20	20	0.876	0.876	0.876
	SVM	17	21	0.892	0.870	0.881
	RF	37	32	0.778	0.802	0.790
	GB	41	30	0.7763	0.814	0.788
	MLP	24	23	0.852	0.858	0.855
Unique attributes	k-NN	27	41	0.817	0.747	0.780
	Logistic	20	21	0.875	0.870	0.873
	SVM	17	21	0.892	0.870	0.881
	RF	32	32	0.802	0.802	0.802
	GB	27	28	0.832	0.827	0.829
	MLP	24	25	0.851	0.845	0.848
Polarity-oriented unique attributes	k-NN	30	43	0.798	0.734	0.765
	Logistic	20	33	0.865	0.796	0.830
	SVM	21	34	0.859	0.790	0.833
	RF	45	36	0.736	0.777	0.756
	GB	41	31	0.761	0.808	0.784
	MLP	24	37	0.839	0.771	0.804

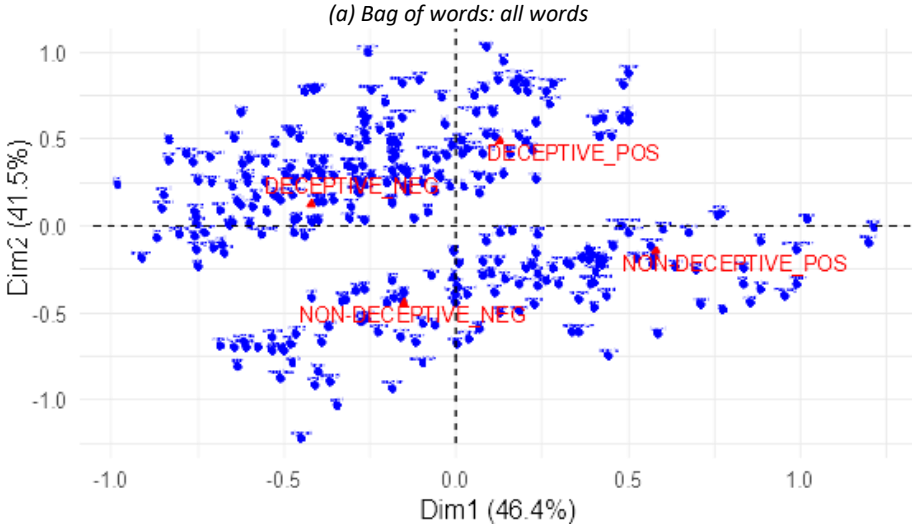
The advantage of using the polarity-oriented unique attributes is illustrated in Figure 3. A correspondence analysis was applied to find the relationships among the set of attributes of

each bag of words and the classes represented by the following four classes: deceptive positively oriented (DECEPTIVE_POS), non-deceptive positively oriented (NON-DECEPTIVE_POS), deceptive negatively oriented (DECEPTIVE_NEG), non-deceptive negatively oriented (NON-DECEPTIVE_NEG). The correspondence analysis mainly utilizes the coordinates on the bi-plot, which is the basic outcome of this analysis. Its shows the correspondence between the items of the two basic categories, classes and attributes, according to their distance to each other (Greenacre, M., & Blasius, 2006). The overall results of correspondence analysis are shown in Table 4, and it includes the number of dimensions, the eigenvalues and the proportions of explained variance from calculated dimensions.

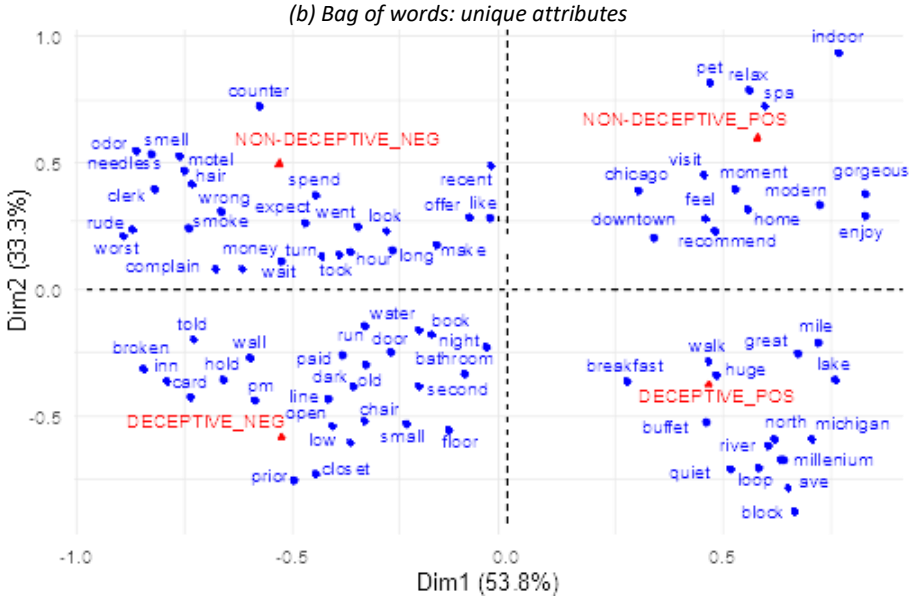
Figure 3 (a) details the correspondence analysis for the case of all words. It can be noticed that there is no clear words correspondence to each one of the four classes, and there is no clear separation. Therefore, using this set of attributes is difficult to define the topics belonging to each class. Figure 3 (b) depicts the case of unique attributes. In this case there is a clear separation of attributes associated to deceptive and to non-deceptive classes. However, they are not separated by the polarity of opinions, which has been claimed as a differentiating feature of deceptive and non-deceptive opinions. Finally, Figure 3 (c) shows the case of polarity-oriented unique attributes, where they are clearly separated by the type of opinions and by their polarity. Therefore, it can be concluded that polarity -oriented attributes improves the association with positive and negative deceptive and non-deceptive reviews. The results detailed in Table 4 shows that the two dimensions represented in the bi-plot account over 85% of the variance, which means that with two dimensions we are able to explain over 85% of the association between classes and attributes. The significance of the association between classes and attributes is given by the chi-square test (F and p-value detailed in Table 4), which means a highly significant association. So, it can be said that the bi-plot of the correspondence analysis explain the relationships between classes and attributes in all cases, although this association exhibit a clear separation in the case of polarity-oriented unique attributes. Therefore, this set of attributed improves the association with positive and negative deceptive and non-deceptive reviews, as established by hypothesis H2.



(a) Bag of words: all words



(b) Bag of words: unique attributes



(c) Bag of words: polarity-oriented unique attributes

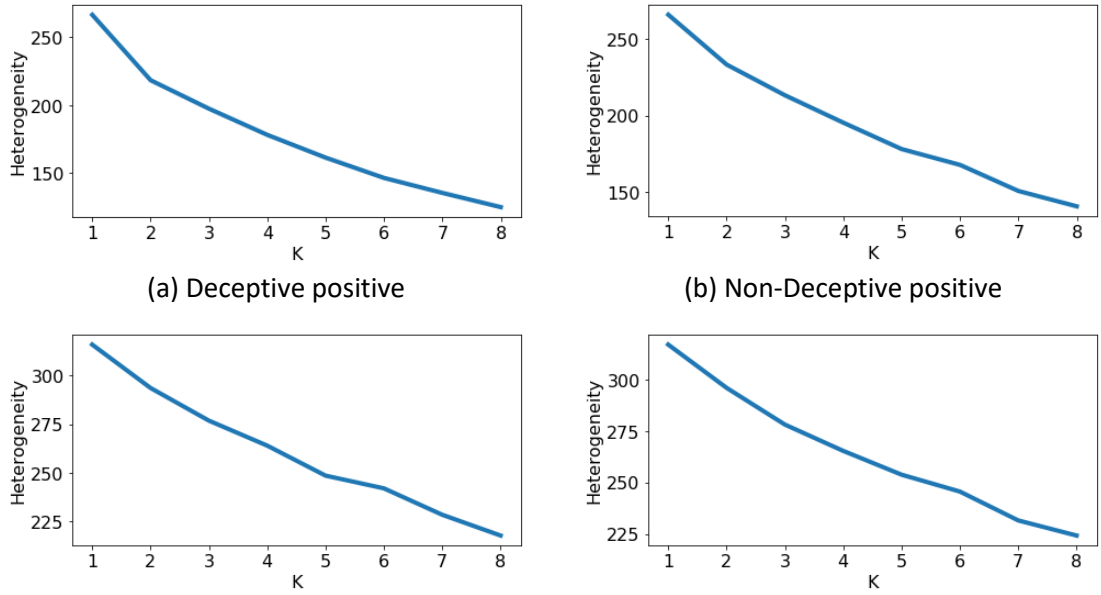
Figure 3. Correspondence analysis.

Table 4. Statistical summary of correspondence analysis between classes and attributes for the three options of bag of words.

Bag of words	Dimension	Eigenvalue	Variance (%)	Cum. Variance (%)
All words	Dim 1	0.139	60.62	60.62
	Dim 2	0.063	27.42	88.04
	Dim 3	0.027	11.95	100.00
	F=21715.49, p-value=0			
Unique attributes	Dim 1	0.132	46.37	46.37
	Dim 2	0.118	41.52	87.89
	Dim 3	0.034	12.10	100.00
	F=11358.77, p-value=0			
Polarity-oriented unique attributes	Dim 1	0.213	54.13	54.13
	Dim 2	0.124	31.71	85.84
	Dim 3	0.056	14.16	100.00

Bag of words	Dimension	Eigenvalue	Variance (%)	Cum. Variance (%)
	F=9230.05, p-value=0			

By using the polarity-oriented unique attributes, we will obtain the distinguishing topics that belong to each of the four classes. To this aim, we apply a clustering algorithm to the attributes with a clear association to each class, so we can compare the topics of positive deceptive and non-deceptive reviews on one hand, and the topics of negative deceptive and non-deceptive reviews on the other. The clustering of documents was performed using the k-means algorithm, and value of *k* was selected using the elbow criterion. Figure 4 depicts the evolution of heterogeneity (sum of squares of the distances to centroids) with the value of k. The elbow criterion considers the optimum value of k as the elbow of the curve.



(a) Deceptive positive (b) Non-Deceptive positive

(c) Deceptive negative (d) Non-Deceptive negative

Figure 4. Selection of the value of k for the k-means clustering algorithm using the elbow criterion.

Following this criterion, we obtain the topics for the four classes given by deception and polarity. Figure 5 depicts discovered clusters over the bi-plot of polarity-oriented unique attributes, which lead to the distinguishing topics among deceptive and non-deceptive positive/negative reviews.

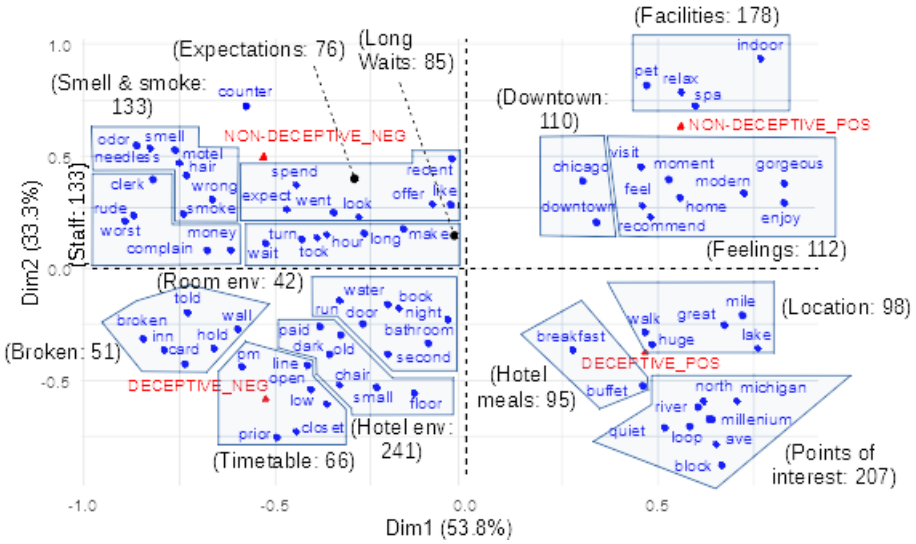


Figure 5. Clusters of topics given by polarity-oriented unique attributes.

Regarding the positive deceptive opinions, the value of k was set to 3 as given by Figure 4 (a). The clustering analysis aggregates the opinions belonging to this class in three topics related to the location of the hotel (98 opinions), the hotel meals (breakfast, buffet, 95 opinions) and the points of interest of the city related to the location of the hotel (207 opinions). We compare these topics to those obtained for positive non-deceptive opinions. In this case, Figure 4 (b) suggest also 3 clusters, and the obtained topics after applying the cluster analysis refers to Chicago downtown (110 opinions), recommendations about visiting Chicago (112 opinions) and hotel facilities (178 opinions). Table 5 details the three topics of positive deceptive and non-deceptive opinions, their unique attributes, and two sentences per topic collected from the dataset.

Table 5. Topics, unique attributes and examples of deceptive and non-deceptive positive opinions.

Topics of deceptive positive opinions	Attributes	Examples
Location	walk, great, mile, huge, lake	"Most major restaurants, Shopping, Sightseeing attractions within walking distance"
		"Our Suite has view on Michigan avenue with a bit of the Lake at the end"
Hotel meals	Breakfast, buffet	"The breakfast buffet was great and was served in the hotels huge atrium which was nice"
		"We ate at their restaurant twice and the breakfast buffet was delicious"
Points of interest	north, michigan, river, millenium, quiet, loop, ave, block	"We stayed in a Parkview Suite so we had a large room with a great view of Millenium Park"
		"Corner tower suite with a view of the river and Michigan ave"

Topics of non-deceptive positive opinions	Attributes	Examples
Chicago downtown	chicago, downtown	"My husband and I would highly recommend this hotel to anyone visiting downtown Chicago"
		"If you want the downtown experience of a lifetime, with historical living that will bring you back to Chicago in the early 1900's look no further"
Feelings about the hotel	visit, moment, feel, modern, gorgeous, recommend, home, enjoy	"The room didn't seem like we were in a hotel, it had the feeling of home"
		"The most memorable part of my stay was looking out at the city after dark and seeing how gorgeous Chicago looks all lit up"
Facilities	pet, relax, spa, indoor	"There were also some additional amenities that we appreciated such as high-speed internet access, a wet bar, an indoor swimming pool and two gorgeous sundecks for relaxing outside"
		"A nice aspect that isn't so common to find is the ability to have small pets stay with you in the hotel

The comparison of topics of deceptive and non-deceptive positive reviews reveals important differences. Deceptive positive reviews typically emphasize the location of the hotel and its ubication with respect to some points of interest of Chicago, such as the Michigan Avenue, the Millenium Park or the river, while positive non-deceptive reviews are focused on characteristic of the city, not necessarily linked to the ubication of the hotel, facilities beyond meals (indoor activities, spa, pets), as posited by H3. Regarding H4, deceptive positive reviews put the focus on hotel meals while non-deceptive positive reviews are more likely to write about the feelings and experiences related to the stay at the hotel. The results obtained can be explained because deceptive reviews are specifically focused on the hotel, while honest reviews talk not only about the hotel but also about the global experience of visiting Chicago.

With respect to negative opinions, we found 4 clusters for deceptive and non-deceptive opinions (Figure 4 (c) and (d)). The topics of negative deceptive opinions are focused on complaints about the hotel environment (241 opinions), the room environment (42 opinions), broken things (51 opinions) and failures to keep the timetable (66 opinions). In the case of non-deceptive negative opinions, the complaints are related to long waits (85 opinions), staff behavior (133 opinions), smell and smoke (133 opinions) and unsatisfied expectations (76 opinions). Table 6 details the four topics of negative deceptive and non-deceptive opinions, their unique attributes, and two examples of sentences per topic collected from the dataset.

Table 6. Topics, unique attributes and examples of deceptive and non-deceptive negative opinions

Topics of deceptive negative opinions	Attributes	Examples

Complaints about the hotel environment	paid, dark, old, chair, small, floor	"The air-condition has a noisy fan/compressor in each room. It starts every 3 to 5 min and the noise will wake you up. This is an old fashion system that needs to be replaced"
		"Rooms are so dark we had to insist they bring lamps so we could see"
Complaints about room environment	water, book, run, door, night, bathroom, second	"THEN we get up the next morning at 7:30 to get ready only to find that we have no running water."
		"In the middle of the night the pipes in our room made a very loud vibrating noise which kept us awake"
Complaints about broken things	broken, told, inn, card, hold, wall	"Our suite had a bathroom sliding door between the bedroom and bathroom which appeared broken, and would not move"
		"Room had a broken phone and broken lightbulb"
Complaints about failures to keep de timetable	pm, line, open, low, closet, prior	"We overheard a lady asking when the pool would open as we were checking in. Though my son was eager to swim as soon as we checked in the pool was still closed at 6:30 pm."
		"We checked in at 7 PM and our room wasn't ready"
Topics of non-deceptive negative opinions	**Attributes**	**Examples**
Complaints about long waits	wait, turn, look, hour, long, make	"When I arrived I had to wait in the lobby for 15 minutes before someone came to the front desk to check me in"
		"Two-hours later, someone came into the room to 'investigate'. When they saw the mess that had been left, they offered a menial apology. Still another two hours later, housekeeping arrived"
Complaints about staff behaviour	clerk, rude, worst, complain	"From the moment we arrived, the staff was belligerent and extremely rude"
		"All the staff at this hotel seemed unhappy and barley even acknowledged any of the guests"
Complaints about smell and smoke	odor, smell, needless, motel, hair, wrong, smoke	"I was lured in by the hotel's pictures showing a fabulous suite. Instead I arrived to find cheap furniture that smelled of old cheese"

		"Although I asked for non-smoking, the room reeked of smoke"
Complaints about unsatisfied expectations	expect, spend, went, look, offer, recent, like, money	"for the kind of money we spent for a weekend here, we were expecting at least a little luxury and special treatment"
		"Meanwhile, the hotel doesn't even offer free wireless-an essential feature for business travelers like me"

In the case of deceptive negative, findings reveal that complaints basically address four areas related to tangible problems of facilities and services: hotel environment, room status, broken elements and failures to keep the timetable and services hours. Conversely, complaints of non-deceptive reviews are focused on the bad experiences resulting from these problems. Reviewers like to show their emotions and feelings about unsatisfied expectations as well as their dissatisfactions with some functional aspects of the hotel, such as long waits, unkindness of staff or hotel smell and smoke. Therefore, and as posited by H5, the orientation of deceptive negative complaints is towards more tangible aspects that non-deceptive complaints, which are more on the side of feelings and expectations.

6. Discussions and implications

Fighting against the manipulation of information in the Internet is a priority for the tourist sector, as it can compromise consumers' confidence on online channels. The main challenge is the ability of fraudsters to resemble the profile and opinions of honest reviewers. Although many websites provide some statistics and reputation about users, they can be easily manipulated through the creation of fake profiles. Recent studies (Jiang et al., 2016) point out the need to address the detection of fake reviews using a multifaceted behavioural information integration approach, considering review and reviewer approaches combined with network approaches. This paper advances in the field of review centric approaches by using several machine learning techniques with the aim of obtaining the unique attributes and topics of deceptive and non-deceptive reviews. The key findings of this study are the identification of polarity-oriented unique attributes able to clearly separate between deceptive and non-deceptive reviews by their polarity orientation and the identification of distinguishing topics for deceptive and non-deceptive reviews, also considering their polarity orientation.

6.1 Theoretical contribution

From a theoretical perspective, this paper advances in the application of text mining techniques to online shared reviews. In general, text mining algorithms rely on the selection of a set of features (terms) which then are mathematically computed using their TF or their TF-IDF value. The performance of classifiers heavily depends on the selection of features, which must have some discriminative properties among the considered classes. Typically, TF-IDF value is a normalized value that emphasize those terms that are common in a subset of documents but not in all of them (in such case, the inverse document frequency is low). However, there is no guarantee that the subset of documents represented by a term with a high TF-IDF value matches

those belonging to one of the pre-defined classes of the problem. The method consisting of selecting unique attributes proposed by Toral et al. (2018) overcomes this problem as attributes are specifically associated to one of the pre-defined classes. This paper introduces the idea of polarity-oriented unique attributes, so the classes of the problem are extended by considering the polarity orientation of reviews. Previous works address that opinions' polarity plays an important role in the detection of deception (Fusilier et al., 2015). The assumption that opinions' polarity is known a priori is justified because it is much easier to detect the sentiment polarity than their truthfulness (Zhang et al., 2018). The bi-plot resulting from the correspondence analysis clearly shows the benefits of selecting this set of attributes instead of unique attributes without polarity orientation, or the case of all words.

As another contribution, the paper also provides a method for obtaining the distinguishing topics of honest and deceptive positive reviews on one hand, and honest and deceptive negative reviews on the other. Previous works about the topics of deceptive reviews were only focused on the deceptive side (Chen et al., 2017). However, and as an advance over these studies, finding unique topics associated to deceptive and non-deceptive reviews can help to prevent fraudulent activities.

6.2 Managerial implications

The findings of our study offer interesting implications for review site operators and hospitality sector. First, review site operators require an automatic system able to detect deceptive reviews prior to its publication, or at least a system that could inform about the trustworthiness of reviews. The need for an automatic system is justified because review sites receive thousands of reviews that cannot be manually checked, as this is a highly time-consuming and cost-intensive human task. The main challenge for such system is that fraudsters try to resemble both the profile and style of normal reviewers to improve the credibility of their shared opinions. By using advanced artificially intelligence algorithms, computer programs and bots can create new reviews with a specific positive or negative purpose. Our findings reveal that it is possible to distinguish deceptive and non-deceptive reviews as they address specific and different topics. For instance, in the case of positive reviews, deceptive ones address common facilities of hotels and well known point of interest of the city, while non-deceptive reviews address more specific facilities that require a prior knowledge of the hotel as well as feelings and emotions linked to the facilities, more difficult to be copied for a manipulated review. Thus, our findings provide new features to be used in conjunction with additional ones to achieve a multifaceted behavioural information integration approach, a pointed out by Jiang et al. (2016).

Regarding the hospitality industry, its scope of action to handle malicious practices is limited. Chevalier & Mayzlin (2006) find that consumer purchasing behaviour responds less intensively to positive reviews (which consumers may estimate are more frequently fake) than to negative reviews (which consumers may assess to be more frequently unbiased). Therefore, the most advisable course of action for hotel managers consists of actively address responses to negative reviews. To increase the effectiveness of such responses, they should be able to discriminate those reviews coming from real reviewers.

6.3 Limitations and further research

There are several limitations of this work. First, the dataset corresponds of truthful and deceptive hotel reviews of 20 most popular Chicago hotels. The data size of 800 reviews was not big enough, which might have introduced some biases in the results. However, it is worth mentioning the difficulty of collecting a manually annotated dataset of deceptive and non-

deceptive reviews. Additionally, although the geographical scope is limited to a city, it is important to know the specific context of the city to extract conclusions about the distinguishing topics resulting from the proposed analysis.

A second limitation is that we follow a review centric approach, so we do not include anything about user profiles or user networking activities. Regarding the content approach followed in this paper, only unigrams (single words) were considered for the analysis. Adding bi-grams (sequence of two adjacent words) could provide an easier interpretation of features. However, the frequency of bi-grams is usually much lower that the frequency of unigrams, so it is not expected that the addition of bi-grams could seriously impact the results.

As a future work, our study could be extended by considering other cities or even other sectors different to the hospitality sector. Moreover, our approach could be combined with other approaches based on reviewers' profiles and their networking activities to build an integrated and more robust system. Also, an independent analysis for different hotel classes (i.e., star ratings) could be conducted.

Additionally, it would be also interesting to conduct a longitudinal study to check how the polarity-oriented unique attributes and the distinguishing topics change over time. The reason is that fraudsters can also change their patterns of action over time. The development of artificial intelligence and deep learning is making possible the artificial creation of reviews by bots or algorithms that learn from honest reviews, making the much harder the identification of deceptive reviews. However, other emotional states such as happiness, sadness, anger, etc. could be considered to identify deceptive reviews

7. Conclusions

Most works in fake reviews detection use and combine different methods to generate features which allow identifying these. However, they do not consider that features may change influenced by the nature of the text. This paper follows a review centric approach and determines the polarity-oriented unique attributes and the polarity-oriented unique topics for deceptive and non-deceptive reviews.

Our study demonstrates that it is possible to distinguish deceptive and non-deceptive reviews in the basis of a set of attributes. This set also facilitates the interpretation of topics uniquely associate to each class of reviews considering their polarity.

Contribution made by each author to the paper

Dr Martínez-Torres as the lead author collected the research data, performed data analysis and drafted the paper. The theoretical and methodological perspectives adopted in the paper were devised in collaboration with Dr Toral, as well as formulation of the main contributions of the research and the discussion section.

Dr Toral assisted with the implementation of machine learning techniques and correspondence analysis.

References

[1] Agnihotri, D., Verma, K., & Tripathi, P. (2017). Variable global feature selection scheme for automatic classification of text documents. Expert Systems with Applications, 81, 268-281.

[2] Ahsan, M. I., Nahian, T., Kafi, A. A., Hossain, M. I., & Shah, F. M. (2016). Review spam detection using active learning. In Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2016 IEEE 7th Annual (pp. 1-7). IEEE.

[3] Akoglu, L., Chandy, R., & Faloutsos, C. (2013). Opinion Fraud Detection in Online Reviews by Network Effects. ICWSM, 13, 2-11.

[4] Akoglu, L., McGlohon, M., & Faloutsos, C. (2010). Oddball: Spotting anomalies in weighted graphs. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 410-421). Springer, Berlin, Heidelberg.

[5] Arenas-Marquez, F. J., Martínez-Torres, M. R., Toral, S., 2014. Electronic word of mouth communities from the perspective of Social Network Analysis. Technology Analysis & Strategic Management 26 (8), 927-942.

[6] Banerjee, S., & Chua, A. (2014). Dissecting genuine and deceptive kudos: The case of online hotel reviews. IJACSA) International Journal of Advanced Computer Science and Applications, 4(3), 2014

[7] Banarjee, S., Bhattacharyya, S., & Bose, I. (2017). Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. Decision Support Systems, 96, 17-26.

[8] Barbado, R., Araque, O., & Iglesias, C. A. (2019). A framework for fake review detection in online consumer electronics retailers. Information Processing & Management, 56(4), 1234-1244.

[9] Barsky, J. D., & Labagh, R. (1992). A strategy for customer satisfaction. Cornell Hotel and Restaurant Administration Quarterly, 33(5), 32-40.

[10] Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. Journal of Hospitality Marketing & Management, 25(1), 1-24.

[11] Bi, J. W., Liu, Y., Fan, Z. P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. Tourism Management, 70, 460-478.

[12] Chen, C., Wen, S., Zhang, J., Xiang, Y., Oliver, J., Alelaiwi, A., & Hassan, M. M. (2017). Investigating the deceptive information in Twitter spam. Future Generation Computer Systems, 72, 319-326.

[13] Cheung, M. Y., Luo, C., Sia, C. L., Chen, H., 2009. Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. International Journal of Electronic Commerce 13 (4), 9-38.

[14] Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. Journal of marketing research, 43(3), 345-354.

[15] Chirita, P. A., Diederich, J., & Nejdl, W. (2005). MailRank: using ranking for spam detection. In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 373-380). ACM.

[16] Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1-4.

[17] Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Al Najada, H. (2015). Survey of review spam detection using machine learning techniques. Journal of Big Data, 2(1), 23.

[18]Do, T. D., Hui, S. C., & Fong, A. C. (2006). Associative feature selection for text mining. International Journal of Information Technology, 12(4), 59-68.
[19]Feng, S., Banerjee, R., & Choi, Y. (2012, July). Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2 (pp. 171-175). Association for Computational Linguistics.
[20]Fusilier, D. H., Montes-y-Gómez, M., Rosso, P., & Cabrera, R. G. (2015). Detecting positive and negative deceptive opinions using PU-learning. Information processing & management, 51(4), 433-443.
[21]Gao, X., Li, S., Zhu, Y., nan, Y., Jian, Z. & Tang, H. (2019). Identification of deceptive reviews by sentimental analysis and characteristics of reviewers. Journal of Engineering Science and Technology Review, 12 (1), 196-202.
[22]Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. Tourism Management, 61, 43-54.
[23]Gonzalez-Rodriguez, M. R., Martínez-Torres, M. R., Toral, S. L., (2016). Post-visit and pre-visit tourist destination image through eWOM sentiment analysis and perceived helpfulness. International Journal of Contemporary Hospitality Management 28 (11).
[24]Greenacre, M., & Blasius, J. (2006). Multiple correspondence analysis and related methods. Chapman and Hall/CRC.
[25]Hancock, J. T., Woodworth, M. T., & Porter, S. (2013). Hungry like the wolf: A word-pattern analysis of the language of psychopaths. Legal and criminological psychology, 18(1), 102-114.
[26]Hernández-Castañeda, A., Calvo, H. & Juárez Gambino, O. (2018). Impact of polarity in deception detection. Journal of Intelligent & Fuzzy Systems, 35, 549-558
[27]Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. Tourism Management, 72, 417-426.
[28]Jiang, M., Cui, P., & Faloutsos, C. (2016). Suspicious behavior detection: Current trends and future directions. IEEE Intelligent Systems, 31(1), 31-39.
[29]Kirilenko, A. P., Stepchenkova, S. O., & Hernandez, J. M. (2019). Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. Tourism Management, 72, 400-410.
[30]Ku, Y.C., Wei, C. P., Hsiao, H. W., 2012. To whom should I listen? Finding reputable reviewers in opinion-sharing communities, Decision Support Systems 3, 534-542.
[31]Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. Journal of Accounting Research, 50(2), 495-540.
[32]Lau, R. Y., Liao, S. Y., Kwok, R. C. W., Xu, K., Xia, Y., & Li, Y. (2011). Text mining and probabilistic language modeling for online review spam detecting. ACM Transactions on Management Information Systems, 2(4), 1-30.
[33]Lee, C. C., & Hu, C. (2005). Analyzing Hotel customers' E-complaints from an internet complaint forum. Journal of Travel & Tourism Marketing, 17(2-3), 167-181.
[34]Li, H., Chen, Z., Liu, B., Wei, X., Shao, J., 2014. Spotting fake reviews via collective positive-unlabeled learning. In 2014 IEEE International Conference on Data Mining, 899-904.
[35]Li, C., Cui, G., & Peng, L. (2017). The signaling effect of management response in engaging customers: A study of the hotel industry. Tourism Management, 62, 42-53.
[36]Luo, G. (2016). A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Network Modeling Analysis in Health Informatics and Bioinformatics, 5(1), 18.
[37]Markowitz, D. M., & Hancock, J. T. (2014). Linguistic traces of a scientific fraud: The case of Diederik Stapel. PloS one, 9(8), e105937.

[38]Martínez-Torres, M. R., Arenas-Marquez, F. J., Olmedilla, M., & Toral, S. L. (2018). Identifying the features of reputable users in eWOM communities by using Particle Swarm Optimization. Technological Forecasting and Social Change, 133, 220-228.

[39]Mankad, S., Han, H. S., Goh, J., & Gavirneni, S. (2016). Understanding online hotel reviews through automated text analysis. *Service Science*, *8*(2), 124-138.

[40]Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. American Economic Review, 104(8), 2421-55.

[41]Olmedilla, M., Martínez-Torres, M. R., & Toral, S. L. (2016a). Harvesting Big Data in social science: A methodological approach for collecting online user-generated content. Computer Standards & Interfaces, 46, 79-87.

[42]Ong, T., Mannino, M., & Gregg, D. (2014). Linguistic characteristics of shill reviews. Electronic Commerce Research and Applications, 13(2), 69-78.

[43]Ott, M., Cardie, C., & Hancock, J. T. (2013). Negative deceptive opinion spam. In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies (pp. 497-501).

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

[45]Parapar, J., Losada, D. E., & Barreiro, A. (2014). Combining Psycho-linguistic, Content-based and Chat-based Features to Detect Predation in Chatrooms. Journal of Universal Computer Science, 20(2), 213-239.

[46]Porter, M. F. (1980). An algorithm for suffix stripping. Program, 14(3), 130-137.

[47]Roberts, D., 2013. Yelp's FakeReviewProblem. ⟨http://tech.fortune.cnn.com/2013/09/26/yelps-fake-review-problem/⟩. Accessed January 25, 2018.

[48]Shan, Y., 2016. How credible are online product reviews? The effects of self-generated and system-generated cues on source credibility evaluation. Computers in Human Behavior 55, 633-641.

[49]Shojaee, S., Murad, M. A. A., Azman, A. B., Sharef, N. M., & Nadali, S. (2013). Detecting deceptive reviews using lexical and syntactic features. In Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on (pp. 53-58). IEEE.

[50]Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. Journal of Travel & Tourism Marketing, 32(5), 608-621..

[51]Teso, E., Olmedilla, M., Martínez-Torres, M. R., & Toral, S. L. (2018). Application of text mining techniques to the analysis of discourse in eWOM communications from a gender perspective. Technological Forecasting and Social Change, 129, 131-142.

[52]Toral, S. L., Martínez-Torres, M. R., & Gonzalez-Rodriguez, M. R. (2018). Identification of the Unique Attributes of Tourist Destinations from Online Reviews. Journal of Travel Research, 57(7), 908-919.

[53]Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), 24-54.

[54]Sun, H., Morales, A., & Yan, X. (2013, August). Synthetic review spamming and defense. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1088-1096). ACM.

[55]Whitlark, D. B., and S. M. Smith. 2001. "Using correspondence analysis to map relationships." Marketing Research, 13(3): 22.

[56]Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. IEEE transactions on evolutionary computation, 1(1), 67-82.
[57]Wu, P. F., 2013. In search of negativity bias: An empirical study of perceived helpfulness of online reviews. Psychology & Marketing 30 (11), 971-984.
[58]Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. International journal of hospitality management, 55, 57-69.
[59]Yee Liau, B., & Pei Tan, P. (2014). Gaining customer knowledge in low cost airlines through text mining. Industrial management & data systems, 114(9), 1344-1359.
[60]Youn Kim, H., and J. H. Yoon. 2013. "Examining national tourism brand image: content analysis of Lonely Planet Korea." Tourism Review, 68(2): 56-71.
[61]Zhang, H., Fan, Z., Zheng, J., & Liu, Q. (2012). An improving deception detection method in computer-mediated communication. Journal of Networks, 7(11), 1811.
[62]Zhang, W., Du, Y., Yoshida, T., & Wang, Q. (2018). DRI-RCNN: An approach to deceptive review identification using recurrent convolutional neural network. Information Processing & Management, 54(4), 576-592.
[63]Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of marketing, 74(2), 133-148.

Prof. María del Rocío Martínez-Torres (rmtorres@us.es) is a full Professor in Management and Business Administration at Business Administration and Marketing Department, University of Seville. Her main research interests include Intellectual Capital and Knowledge Management, Social Network Analysis, Open Innovation and Virtual Communities

Dr. Sergio Toral (storal@us.es) is a full Professor in Digital Electronic Systems at the Department of Electronic Engineering, University of Seville. His main research interests include Open Source Software projects, Open Innovation and Social Network Analysis