



HAL
open science

Deceptive Opinions Detection Using New Proposed Arabic Semantic Features

Amel Ziani, Nabiha Azizi, Didier Schwab, Djamel Zenakhra, Monther Aldwairi, Nassira Chekkai, Nawel Zemmal, Marwa Hadj Salah

► **To cite this version:**

Amel Ziani, Nabiha Azizi, Didier Schwab, Djamel Zenakhra, Monther Aldwairi, et al.. Deceptive Opinions Detection Using New Proposed Arabic Semantic Features. *Procedia Computer Science*, 2021, 189, pp.29 - 36. 10.1016/j.procs.2021.05.067 . hal-03299022

HAL Id: hal-03299022

<https://hal.science/hal-03299022>

Submitted on 25 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Deceptive Opinions Detection Using New Proposed Arabic Semantic Features

Amel Ziani ^{a,b*}, Nabihaz Azizi^c, Didier Schwab^d, Djamel Zenakhra^c, Monther Aldwairi^e,
Nassira Chekkai^f, Nawel Zemmal^{c,g}, Marwa Hadj Salah^{d,h}

^a*Benyoucef Benkhedda alger 1 university, Algiers, Algeria.*

^b*LRI: Information Retrieval Laboratory, Badji Mokhtar University, Annaba, Algeria*

^c*LABGED: Electronic documents control laboratory, Badji Mokhtar University, Annaba, Algeria*

^d*LIG-GETALP Univ. Grenoble Alpes, France*

^e*College of Technological Innovation, Zayed University, P.O. Box 144534 Abu Dhabi, United Arab Emirates*

^f*Alger 3 daly Ibrahim university, Algiers, Algeria.*

^g*Mohamed-Cherif Messaadia University, Souk Ahras, Algeria*

^h*LATICE, Faculty of Economics and Management of Sfax, Sfax, Tunisia*

Abstract

Some users try to post false reviews to promote or to devalue other's products and services. This action is known as deceptive opinions spam, where spammers try to gain or to profit from posting untruthful reviews. Therefore, we conducted this work to develop and to implement new semantic features to improve the Arabic deception detection. These features were inspired from the study of discourse parse and the rhetoric relations in Arabic. Looking to the importance of the phrase unit in the Arabic language and the grammatical studies, we have analyzed and selected the most used unit markers and relations to calculate the proposed features. These last were used basically to represent the reviews texts in the classification phase. Thus, the most accurate classification technique used in this area which has been proven by several previous works is the Support Vector Machine classifier (SVM). But there is always a lack concerning the Arabic annotated resources specially for deception detection area as it is considered new research area. Therefore, we used the semi supervised SVM to overcome this problem by using the unlabeled data.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 5th International Conference on AI in Computational Linguistics.

Keywords: Deceptive Opinions Detection; Opinion Mining; Arabic Language; Semantic Features; Support Vector Machine; Semi Supervised Learning;

* Corresponding author. Tel.: +213 556678594.

E-mail address: z_amel1911@live.fr/zianiamel1911@gmail.com

1. Introduction

Online reviews are becoming the cornerstone for users' decisions on the ever-growing social media networks, online shopping and blogs. However, sometimes these reviews may be misleading. Deceptive opinion detection is a critical and important task for opinions analysis and recommendation systems. Where the main goal of this task is to eliminate the suspicious, fake or trolling opinions leaving only the genuine and organic ones ([1]; [2]; [3]; [4]; [5]; [6]; [7]; [8]; [9]; [10]; [11]). Deceptive opinion is the review with trickery or fake opinions, designed to misguide the users or to attract people's attention. For example, to promote or damage the reputation of products ([12]; [13]). Deceptive opinions detection can be considered as bi-class classification problem. Most studies use machine learning approaches for the classification phase, such as Ott et al. (2011). Therefore, to enhance the classification performance most works focused on finding the most effective features. Feature engineering is important; however, it is difficult to represent data with statistically, syntactically or semantically accurate characteristics. As most previous works studied the syntactic features extraction, they were based on analyzing local information, but none of them have explored semantic features for Arabic deceptive opinions detection. That is attributed to the exceptional structure of Arabic phrases. Therefore, we have devoted our latest researches to discover deeper semantic features for Arabic deceptive opinions detection.

This work is primarily concerned with the task of opinions classification, this process deals with an important phase, the extraction of the features vector. Representing a review text with features vector is considered to be most efficient way in processing, even for Arabic datasets, because of its complex morphology [14]. Therefore, we used a set of lexical features chosen and tested in few earlier works ([15]; [16]), namely: emotionalism, reflexivity, addressing, number of positive words, etc. These features were used in multiple experiments and have proven their effectiveness in different propositions, but there was always a weakness in the reviews representation that influenced the system performance. The inadequate use of statistical and semantic features for representing the reviews in an efficient way, led us to explore new features with semantic aspects for better representation of the Arabic texts. Therefore, the idea of the proposed approach is to fuse the two types of features; the old lexical set and the new semantic set to profit from all the aspects of Arabic phrase structure. Thus, we explore the possibility of integrating the polarity of phrase connectors generated by a discourse parser into the process of extracting the features in order to detect the polarity of deceptive opinions. This proposition has been supported by the important effect of discourse processing and coherence relations in the opinion analysis. Where, the general meaning of an opinion or its polarity can be affected by the relation between its phrases, as any phrase polarity can be reinforced by the polarity of their successor or ancestor or can be a negation. All these factors have been taken into account in the feature's extraction phase for the classification of deceptive opinions. SVM classifiers have showed their effectiveness in opinion mining as supervised learning algorithm. In a previous work, we have carried out an experiment to test the performance of multiple classifiers such as; Support Vector Machine, Multi-Layer Perceptron, Naïve Bayes, Multinomial Naïve Bayes, Random Sub Space and Decision Trees. That experiment has proven that there exist other classifiers, other than SVM that gives higher rates for opinion mining such as Multi-Layer Perceptron. Nevertheless, SVM is considered to be more flexible and easier to adapt with other techniques without losing its advantages, this was concluded during another previous study, that analyzed the SVM classifier and its entire kernel functions. Noting that the most of available datasets for deceptive detection are semi labeled and none of them are in Arabic language. And for the goal of considering all the datasets instances (the labeled and the unlabeled) to ameliorate the separation margin between the classes, we have been encouraged to use the semi-supervised (S3VM).

In the remainder of this paper, firstly we outline and detail the proposed approach, with all the phases; the data acquisition, the features extraction, the deceptive opinions detection with S3VM. Later, we discuss the experimental results.

2. Structure Proposed system for the Arabic opinions polarity detection

The internet users have always been interested of knowing other's opinions and recommendations. In order to satisfy their needs, an Arabic deceptive opinions detection system has been proposed in this paper to detect the deception in the online reviews. Nevertheless, to be able to design the system architecture, we realized that this architecture should meet the following requirements (Fig.1):

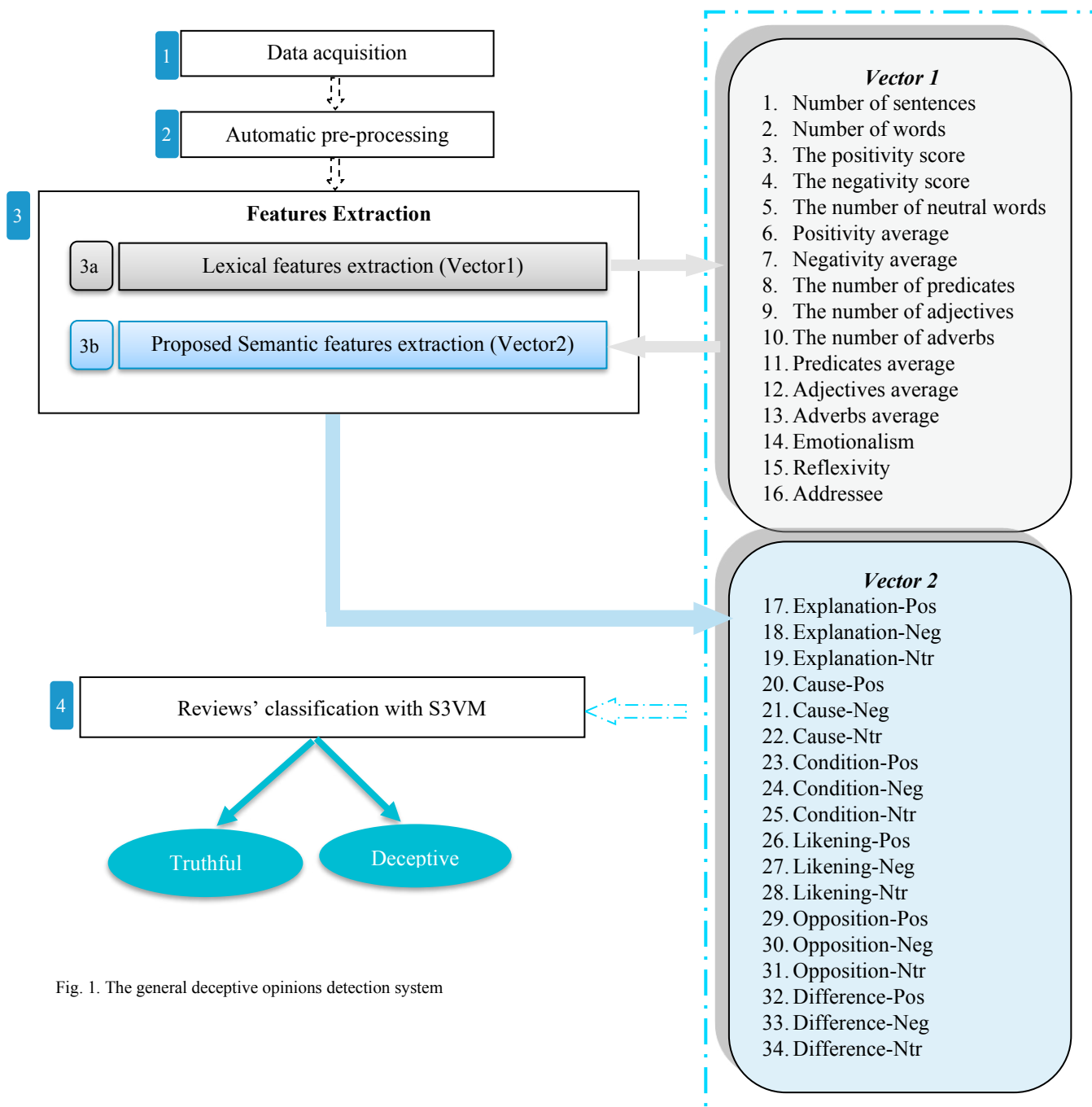


Fig. 1. The general deceptive opinions detection system

2.1. Data acquisition

The task of obtaining annotated data for deceptive opinions detection is becoming the biggest concern for researchers. In [17], the authors used Dianping dataset that consists of reviews of popular restaurants in Shanghai, China from Nov. 2011 to Apr. 2014. Each review is labeled as spam or non-spam using Dianping's commercial spam filter. They take a reviewer as a spammer if s/he has at least 10% of his/her reviews detected as fake/spam by Dianping. They proved that this cutoff allows for some errors in Dianping's detection. Also, among the reviewers with at least one spam review, only 2:3% of them have less than 10% spam reviews.

Other researchers ([19]; [12]; [20]) reported analyses of the Yelp filter based on reviews they crawled. They assumed those reviews which are filtered by Yelp are spam and compiled two datasets respectively: Yelp-Chicago [19] and YelpZip [20]. However, these datasets do not have all reviews of each reviewer as they crawled Yelp reviews based on products. Ott et al. (2011) have proposed an approach for generating positive deceptive opinion spam using Amazon’s popular Mechanical Turk crowdsourcing service [12]. Later Ott et al. (2013) used Amazon’s Mechanical Turk service to produce another publicly available dataset of negative deceptive opinion spam [18].

Therefore, there is no suitable Arabic dataset for our experiments which require complete well-structured reviews in Arabic. Thus, and after we analyzed the mentioned English dataset we noticed that they generated a considerable number of structured truthful and deceptive opinions in English, what made us think in translate it to Arabic to construct an Arabic dataset for our experiments. We used this solution to overcome the lack of Arabic deception datasets and to use meaningful structured texts.

2.2. Feature extraction

The feature extraction is the process of extracting the main characteristics of the text. For a machine learning algorithm to perform well, it is essential to have features that are descriptive of the text.

The common classifiers and learning algorithms cannot handle the emotional text directly. So, we have to represent them in the form that classification algorithm can deal with. The documents are typically represented by feature vector.

In this section, we describe in detail the list of lexical and semantic features we have constructed for the classification task.

a) Lexical features

The selected statistical features were used and tested in many previous studies, that is why we were able to analyze them through all the experiment and notice that despite their effectiveness with small data it doesn’t represent well the texts in Arabic because the Arabic language has a very complex morphology and have special relations. That is what makes those relations important and can influence the general polarity of an opinion. And that what inspired us to propose new semantic features based on rhetoric relations.

b) The proposed semantic features extraction

By analyzing different studies, we noticed the important effect of discourse processing and coherence relations in the opinion analysis. Where, the general meaning of an opinion or its polarity can be affected by the relation. So, it can reinforce it or contradict it as it is explained in the following examples:

أنا أحببت جدا المقال لأنه وصف المباراة بالتفصيل بالرغم من أن المباراة كانت فاشلة وخسر فريقنا ولكننا نشكر سيدي الكاتب على مجهودك.

The coherence relations (Discourse connectives): لأنه، بالرغم، و لكننا

As the second EDU’s polarity is positive it reinforces the polarity of first EDU because of the explanation relation that connects between these two EDUs.

We can conclude that the polarity score of the relation can affect the general polarity score of the phrase, which can be: positive, negative or neutral. Therefore, we have proposed a new set of features that helps in representing the opinion text. So, our proposition is based on the discourse relations, as mentioned before there are too many to work with but as the Arabic language is very rich with connections letters (horouf), we chose only the most important ones in Arabic discourse, which are: explanation, cause, condition, likening, opposition and difference. The proposed features are shown in table 1:

Table 1. The proposed features.

1	Explanation_Pos	If the classification result of both the explanation EDUs is positive; Explanation_Pos= Explanation_Pos +1
---	-----------------	---

2	Explanation_Neg	If the classification result of both the explanation EDUs is negative; Explanation_Neg= Explanation_Neg +1
3	Explanation_Ntr	If the classification result of both the explanation EDUs is neutral; Explanation_Ntr= Explanation_Ntr +1
4	Cause_Pos	If the classification result of sub tree the Cause EDUs is positive; Cause_Pos= Cause_Pos +1
5	Cause_Neg	If the classification result of sub tree the Cause EDUs is negative; Cause_Neg= Cause_Neg +1
6	Cause_Ntr	If the classification result of sub tree the Cause EDUs is neutral; Cause_Ntr= Cause_Ntr +1
7	Condition_Pos	If the classification result of sub tree the condition EDUs is positive; Condition_Pos= Condition_Pos +1
8	Condition_Neg	If the classification result of sub tree the condition EDUs is negative; Condition_Neg= Condition_Neg +1
9	Condition_Ntr	If the classification result of sub tree the condition EDUs is neutral; Condition_Ntr= Condition_Ntr +1
10	Likening_Pos	If the classification result of sub tree the likening EDUs is positive; Likening_Pos= Likening_Pos +1
11	Likening_Neg	If the classification result of sub tree the likening EDUs is negative; Likening_Neg= Likening_Neg +1
12	Likening_Ntr	If the classification result of sub tree the likening EDUs is neutral; Likening_Ntr= Likening_Ntr +1
13	Opposition_Pos	If the classification result of sub tree the opposition EDUs is positive; Opposition_Pos= Opposition_Pos +1
14	Opposition_Neg	If the classification result of sub tree the opposition EDUs is negative, Opposition_Neg= Opposition_Neg +1
15	Opposition_Ntr	If the classification result of sub tree the opposition EDUs is neutral, Opposition_Ntr= Opposition_Ntr +1
16	Difference_Pos	If the classification result of sub tree the difference EDUs is positive, Difference_Pos= Difference_Pos +1
17	Difference_Neg	If the classification result of sub tree the difference EDUs is negative, Difference_Neg= Difference_Neg +1
18	Difference_Ntr	If the classification result of sub tree the difference EDUs is neutral, Difference_Ntr=Difference_Ntr +1

In order to calculate the previous features, we follow the following steps.

Step1: Construct the dependency tree.

Step2: Classify the EDUs and calculate the features while parsing the tree with algorithm 2.

Step 3: Get the new features.

Below we discuss each step in more details.

Step 1: Construct the dependency tree

In order to construct the dependency tree, we have to segment the discourse. As we mentioned before in section, the discourse segmentation aims at splitting texts into Elementary Discourse Units (EDUs), which are the sub-phrases that are connected with a coherent relation. That means we need to identify the coherent relations by extracting the existed connectors from table mentioned in section. Using these extracted units, we can construct the tree where:

1. Every EDU is a leaf node in the tree.
2. For every two connected EDUs, there exist a root node containing the connector.
3. To decide what node is the tree head node, we follow an algorithm that uses the priority of the relations. The priority criteria were defined before in our algorithms after a deep analysis of the connectors or “el horouf” in Arabic language where some of these relations have more importance than the other in the discourse. Therefore, the used coherence relations that follow this priority order: Opposition, Difference, Condition, Explanation, Cause, Likening.

After deciding all the units and the tree head, we will be able to construct the tree and to assign to each node and leaf two parameters (Figure 2): class and polarity scores, where these two parameters will be used and calculated in step two.

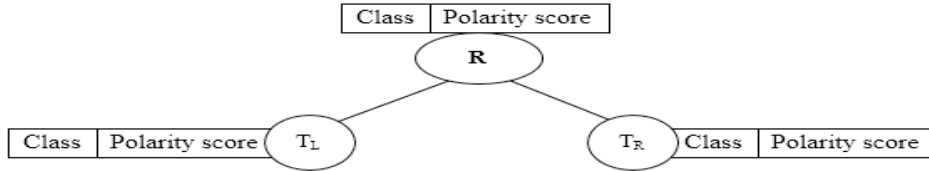


Figure 2. The parameters of each node.

Based on the classification results associated with this tree, we can calculate the new features by detecting each node polarity score.

Step 2: Classification

According to the phrase dependency structure, we parse it hierarchically from leaf nodes to root node. The leaves will be the EDUs of the review text, that's why we have proposed to classify the leaves in order to decide the polarity score and the class of the relation in the parent node.

For each hierarchy in the form of branching triplets (Troot → Tleft Tright), we have to classify each leaf in the tree to decide the root polarity. That will be done by classifying all the leaves using the next algorithm:

The main steps of the classification algorithm:

Step 1: Parse the tree (Left, Right, Root)

Step 2: For each couple leafs (T_{left} , T_{right}), classify sub-phrases (left EDU and right EDU) using the S3VM

If class $T_{left} \neq$ class T_{right} then

The root polarity score = Max polarity score (T_{left} , T_{right})

The root class = the class of the Max polarity score (T_{left} , T_{right})

Else class = class T_{right} and The root polarity score = Max polarity score (T_{left} , T_{right})

Step 3:

For each root node:

Detect the coherent relation and increment the features based on the polarity of the root node.

End.

Our goal from this phase is to calculate the proposed features, which are the relations between the EDUs, where each relation can have a polarity. As we set the root node is the relation, then its polarity or the root polarity is calculated from the leaves of this node.

Therefore, the features are incremented when their associated relation are found in the sub trees:

For example, in case of the root node= explanation then

If the root class = positive, then Explanation_Pos= Explanation_Pos+1;

Else if the root class = negative then Explanation_Neg= Explanation_Neg+1;

We follow the same steps to calculate all the features to construct the new features vector for classification phase.

2.3. Deceptive opinions detection

Supervised learning applicable to deceptive opinions detection, which can be naturally formulated as a classification problem with two classes, deceptive and truthful. However, as mentioned earlier, the key difficulty is that it is very hard, if not impossible, to recognize deceptive reviews reliably by manually reading them because a spammer can carefully craft a fake review that is similar any genuine review [13]. Despite these difficulties, several supervised detection algorithms have been proposed and evaluated in various ways. Only few available datasets provide deceptive and truthful reviews, that is the reason why we proposed to use a semi-supervised algorithm S3VM in order to overcome the lack of annotated data. This classification phase is considered as a multiclass problem, where the resulted classes are (deceptive positive, deceptive negative, truthful positive and truthful negative)

As the name suggests, semi-supervised learning is somewhere between unsupervised and supervised learning. In fact, most semi-supervised learning strategies are based on extending either unsupervised or supervised learning to include additional information typical of the other learning paradigm. Semi-supervised learning is attractive because it can potentially utilize both labeled and unlabeled data to achieve better performance than supervised learning. From a different perspective, semi-supervised learning may achieve the same level of performance as supervised learning, but with fewer labeled instances. This reduces the annotation effort, which leads to reduced cost.

Now to train and classify this semi-supervised classifier we need a perfect representation for the text reviews. For this reason, we have used two sets of features: statistical and semantic.

3. Experimental results and evaluation

As the deception detection field is basically an opinion mining problem, and as our work line is focused on opinions polarity detection, we have carried multiple experiments in order to highlight the importance and to clarify the effect of the proposed semantic features for both cases.

From the various research publications being mentioned before, only the similar works to our proposition were chosen to state their results as they were originally published by authors in order to compare our obtained results to them. The comparison is shown in table 2.

To position our work between the existed works we have made a comparative study with our previous works and some related works with available results.

Table 2. The results of the experiment

Work	Dataset	Features	Language	Precision	Recall	Fmeasure	Accuracy
Previous work	Manual constructed (150 reviews)	Statistical features	Arabic	81.85	85.67	84.76	85.99%
Ziani et al., 2016, [16]	Ott et al. (2013)'s translated dataset (1600)	Statistical features	Arabic	89.10	89.30	87.50	89.33%
Ott et al., 2011, [12]	Ott et al. (2011)'s dataset ()	/	English	89.80	89.80	89.80	89.80%
Ott et al., 2013, [18]	Ott et al. (2013)'s dataset (1600)	/	English	89.10	89.30	88.50	88.40%
Li et al., 2017, [17]	Dianping	/	Chinese	90.00	89.00	85.00	Over 85.0%
Our proposed approach	Ott et al. (2013)'s translated dataset (1600)	Proposed Semantic features and statistical features	Arabic	88.00	96.00	86.00	93%

Based on the obtained experimental results from table 2, we can conclude that the proposed features have shown to be highly effective in Arabic opinion mining outperforming the results obtained by using only statistical features that have been evaluated in previous works. Despite the use of different datasets, the first is for opinion mining and the second for deception detection, we tried to compare with other works and we also obtained better results. This experiment was conducted to prove the importance of using the semantic aspect in Arabic opinion mining.

4. Conclusion

Being able to determine a deceptive opinion from a truthful one is a serious problem in opinion mining and spam detection. The idea of this research is to approach to the right solution for a perfect Arabic deceptive opinions detection system. By working on this, we were able to prove that dealing with Arabic language is much difficult and opposes many morphological and semantic problems that can affect any classification system performance. And looking at the important role of features in any classification system, we focused on exploring and proposing new set of Arabic semantic features that were inspired from the rhetoric phrase dependency algorithms. These last are founded on phrase relations and discourse parsing. Thus, we tried to profit from the deep analysis of the Arabic phrases and the relations polarity in order to calculate the new features. This task was implemented in collaboration with the semi supervised SVM (S3VM) to fulfil the lack of annotated Arabic resources for deception detection. However, this combination and collaboration has been proved to be of a great help to any classification system.

The enhancement made on the proposed approach allowed us to increase the accuracy of the system to 85.99%. There are different directions for extending this system; on one hand, we could improve the Arabic dataset by constructing an annotation system based on active learning technique. In the other hand, it would be worthwhile to integrate the proposed approaches in a recommendation system.

References

- [1] Miller. C. (2009). Company Settles Case of Reviews it Faked, New York Times.
- [2] Meyer. D. (2009). Fake Reviews Prompt Belkin Apology, CNet News.
- [3] Streitfeld. D. (2012). For 2a Star, an Online Retailer Gets 5 Star Product Reviews, 26, New York Times.
- [4] Topping. A. (2010). Historian Orlando Figes Agrees to Pay Damages for Fake Reviews, 16, The Guardian.
- [5] Dhawan. S, and Simran. (2018). An enhanced mechanism of spam and category detection using Neuro-SVM. *Procedia Computer Science*, 132, 429–436.
- [6] Saeed. R. M. K, Rady. S, and Gharib. T. F. (2019). An ensemble approach for spam detection in Arabic opinion texts. *Journal of King Saud University - Computer and Information Sciences*.
- [7] Bajaj. S, Garg. N, and Singh. S. K. (2017). A Novel User-based Spam Review Detection. *Procedia Computer Science*, 122, 1009–1015.
- [8] Chakraborty. M., Pal. S, Pramanik. R, and Ravindranath Chowdary. C. (2016). Recent developments in social spam detection and combating techniques: A survey. *Information Processing and Management*, 52(6), 1053–1073.
- [9] Chen. C, Wen. S, Zhang. J, Xiang. Y, Oliver. J, Alelaiwi. A, and Hassan. M. M. (2016). Investigating the Deceptive Information in Twitter Spam. *Future Generation Computer Systems*.
- [10] Fei. G, Li. H, and Liu. B. (2017). Opinion Spam Detection in Social Networks. *Sentiment Analysis in Social Networks*, 141–156.
- [11] You. L, Peng. Q, Xiong. Z, He. D, Qiu. M, and Zhang. X. (2020). Integrating aspect analysis and local outlier factor for intelligent review spam detection. *Future Generation Computer Systems*, 102, 163–172.
- [12] Ott. M, Choi. Y, Cardie. C, and Hancock. J.T. (2011). Finding deceptive opinion spam by any stretch of the imagination, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics Human Language Technologies*, 1, Association for Computational Linguistics, pp.309–319.
- [13] Jindal. N, Liu. B. (2008). Opinion spam and analysis, in: *Proceedings of the 2008 International Conference on Web Search and Data Mining*, ACM, pp.219–230
- [14] Shaalan. K, Nizar. Y, Habash. (2010). Introduction to Arabic natural language processing (Synthesis lectures on human language technologies). In *Machine translation journal*, pp.285–289.
- [15] Ziani. A, Azizi. N, Tlili. GY. (2015). Combining Random Sub Space Algorithm and Support Vector Machines Classifier for Arabic Opinions Analysis. In *ICCSAMA 2015*, Springer.
- [16] Ziani. A, Kounta. C.A.K, Azizi. N, Zenakhra. D, Cheriguene. S. (2016). A New Hybrid Classification System for Subjectivity Detection. In the proceeding of the CSA Conference on Computing Systems and Applications, December 13-14, 2016, Algiers, Algeria. <http://www.emp.edu.dz/csa/pdf/CSA2016-Proceedings.pdf>
- [17] Li. L, Qin. B, Ren. W, and Liu. T. (2017). Document representation and feature combination for deceptive spam review detection. In *neurocomputing journal*, pp 1339-1351.
- [18] Ott. M, Cardie. C, and Hancock. J. T. (2013). Negative Deceptive Opinion Spam, In *Proceedings of NAACLHLT*, (pp. 497-501).
- [19] Mukherjee. A, Venkataraman. V, Liu. B, and Glance. N. S. (2013). What yelp fake review filter might be doing? In *ICWSM*.
- [20] Rayana. S, and Akoglu. L. (2015). Collective opinion spam detection: Bridging review networks and metadata. In *KDD*, pages 985–994.