# Detection of deceptions in Twitter and News Headlines written in Arabic<sup>\*</sup>

Francisco Eros Blázquez del Rio<sup>1</sup>, Manuel Conde Rodríguez<sup>2</sup> and Jose M. Escalante<sup>3</sup>

<sup>1</sup>frablade77@gmail.com <sup>2</sup>mcondrod@gmail.com <sup>3</sup>jmescalantefernadez@gmail.com

**Abstract.** In this work we present an model to detect deceptive texts (twitters and news headlines) written in Arabic. To develop the model we have focused on the several characteristics to differentiate them from the regular texts, such as: numbers, special characters or n-grams, which could be a new type of deception indicator. The used classifier method is Support Vector Machine, due to it has excellent performance in different low and high-dimensional classification tasks.

Keywords: Support Vector Machine  $\cdot$  Stop of Words  $\cdot$  N-grams  $\cdot$  Special Characters

### Introduction

Currently, the amount of information exchanged by users on the network has reached dimensions that at the beginning of century were considered unimaginable. There are more than 2.5 quintillion bytes of data created every day[1]. Over the last two years alone 90 percent of the data in the world was generated [1]. A large amount of the data generated is due to the explosion of social networks, such as Facebook, Twitter or Instagram [2], which has been favored by the exponential increase in internet speed in the last decade [3] together with the development of wireless communication technology [4]. Among the jungle of different social networks, Twitter (microblogging network of just 140 characters) has been the fastest expanding social network considering its simplicity. Its strength lies in its instantaneity to access information [5]. Due to the instantaneity of this social network, it has a high vulnerability to the propagation of hoaxes, opinions of doubtful credibility and deceptive news.

Although currently we have in the era of information. The misinformation floods our daily life [6,7]. This fact makes mandatory the verification of the veracity of a news or comments posted. The ability of these deceptions (hoaxes and fake news) to shape people's opinion is somewhat worrisome [8], doing that

 $<sup>^{*}\</sup>mathrm{Copyright}$  © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). FIRE 2019, 12-15 December 2019, Kolkata, India.

#### 2 F. E. Blázquez, M. Conde and J. M. Escalante

the propagation of hoax on Twitter and the fake news made are highly correlated due to the characteristic of this social network mentioned above.

There is an increasing interest in deception detection in the context of intelligenceservices, law enforcement or intra-organizational monitoring, but also from a civil point view [9]. Though it is expected that human capability to indulge in deception would be balanced by the ability to detect deception, it has been borne out in studies [10,11,12] that humans are not naturally good in deception detection. In fact, even trained personnel correctly identify deception at rates only a little better than chance (52 percent accuracy) [11].

In contrast, software systems based on number-crunching algorithms and correlation analysis often provide surprisingly accurate results. This is especially true in real-time applications where computerized systems can quickly and efficiently parse huge datasets, flagging objects of interest that can then be sent for final analysis and scrutiny by human judges.

Psychological studies [12,13,14] in the area of deception detection have shown that changes in behaviour such as the body posture, facial expressions, speech rhythm and pitch are closely correlated with deception. Markers of deception that are under conscious control (like the content of the deceptive 'story') may be modified by those who wish to deceive, but fortunately most of these cues are a result of sub conscious processes and therefore even awareness of vigilance and scrutiny does not make deception any easier.

In the context of deception in text, though communication is stripped to its essentials and no non-verbal deception-cues are transmitted, the linguistic manifestations of deception remain consistent across most domains. It has been empirically shown [13,11,15] that deception leaves a linguistic signature caused largely due to the high demands that indulging in deception generates on a person's cognitive capabilities.

Many research groups in the field of psychology [16,15,17] have constructed sets of linguistic markers for deception detection in text using different methodologies. James Pennebaker et al. at the Department of Psychology, University of Texas [14,18] have constructed an empirical deception model based on cue-word usage-frequency profiles. Though this linguistic signature of deception is not easy for humans to detect directly, it is easily detected by software. According to the model, deception in text is marked by:

- Decreased frequency of first-person pronouns (I, mine, myself etc.) a subconscious attempt by the author to disassociate from the deceptive content.
- Decreased frequency of exclusive words (or, but, without etc.) to keep the content simple, concrete and without abstractions in order to avoid faltering while being repeatedly interrogated.
- Increased frequency of negative emotion words (anger, abandon, hate etc.) a reflection of the subconscious feeling of guilt involved with being deceptive.
- Increased frequency of action verbs (move, run, lead etc.) as a form of distraction to keep the 'story' moving while the basic content remains simple and insignificant.

In this work, we try another approach much simpler. The kind of texts considered (twitters and news headlines) have other special characteristics that differentiate them from a regular text. Generally, in this kind of texts it is used numbers, special character or combination of words (n-grams, 2-grams and 4grams) to capture the attention of readers, so the study will be focused in this kind of structures as main signatures of deception[19].

#### Support Vector Machines model

Support Vector Machine (SVM) model is a supervised learning approach introduced by Vapnik in 1995 for solving two-class pattern recognition problem [20]. It is based on the Structural Risk Minimization principle for which error-bound analysis has been theoretically motivated by the works [20,21]. The method is defined over a vector space where the problem is to find a decision surface that best separates the data points in two classes. This model has a excellent performance in different classification task [22,23] comparing with other methods, such as: k-Neares Neighbor (k-NN), Neural Networks (NNet), Linear Leas-Square Fit (LLSF) or Naive Bayes (NV). Figure a shows the training ans test schema of the model.

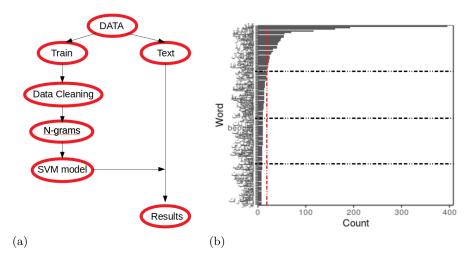


Fig. 1. a) Diagram of training a test os SVM model and b) Frequencies his ogram of BoW (n=100.)

### Methodology

The code, written in R-code initially generates a bag of words (BoW) (n = 1000, number of BoW) and uses a cross validation method (k = 3 and r = 1, where)

4

k and r are parameter to control the cross validation method performance of R-code package) to train a SVM model (method="svmlinear").

After several computation with different values of n, k and r, the chosen values were: n = 100, k = 10 and r = 5, achieving a balance between computation time necessary and the accuracy obtained. Although the reduction in the number of BoW is important, the accuracy is not strongly affected, going from a accuracy of 0.73 to 0.62, but with an important reduction of computation time.

It could be thought that the length of BoW is too short, but taking into account the classes of texts considered (no more 140 characters in the longest case) with n = 100 it is enough. This is supported by Fig. 1b, where it is plotted an histogram the frequencies of every word of BoW. We can see that approximately the first 25 words (n=25) concentrate the highest frequencies of appearance in the texts considered.

The methodology followed is as follows:

- Initially, we train the two variant of SVM model: SVMlinear and SVMlinear3 [26], with the BoW generated.
- 2. Next, we train the two models with a new BoW where the numbers, special character and Stop Words <sup>1</sup> have been removed.
- 3. Next, we train the two models with the N-grams(2 and 4-grams) generated using the texts without removing numbers, special character and Stop Words.
- 4. Finally, the models are trained with the N-grams generated after removing numbers, special character and Stop Words.

The results of accuracies obtained are summarized on the Tables 1 (Twitter) and 2 (News Headlines).

Simulation	BoW	Numbers	Special Character	Stop Words	N-grams $(n=2 \text{ and } 4)$	Model	Accuracy
1	YES	NO	NO	NO	NO	SVMLinear	0.6207
2	YES	NO	NO	NO	NO	SVMLinear3	0.6551
3	YES	YES	YES	YES	NO	SVMLinear	0.6338
4	YES	YES	YES	YES	NO	SVMLinear3	0.6737
5	YES	NO	NO	NO	YES	SVMLinear	0.6319
6	YES	NO	NO	NO	YES	SVMLinear3	0.6564
7	YES	YES	YES	YES	YES	SVMLinear	0.6289
8	YES	YES	YES	YES	YES	SVMLinear3	0.6655

Table 1. Twitter

<sup>&</sup>lt;sup>1</sup>A set of Stop Words is any set of words can be chosen as the stop words for a given purpose [24]. In our case this set of words does not provide any relevant information to decide if a text is true or false. In this work we have use a library of RStudio called "**arabicStemR**" [25] to detect and remove this words.

Table 2. Table 1: News H
--------------------------

Simulation	BoW	Numbers	Special Character	Stop Words	N-grams $(n=2 \text{ and } 4)$	Model	Accuracy
9	YES	NO	NO	NO	NO	SVMLinear	0.6272
10	YES	NO	NO	NO	NO	SVMLinear3	0.6284
11	YES	YES	YES	YES	NO	SVMLinear	0.6165
12	YES	YES	YES	YES	NO	SVMLinear3	0.6202
13	YES	NO	NO	NO	YES	SVMLinear	0.6299
14	YES	NO	NO	NO	YES	SVMLinear3	0.6377
15	YES	YES	YES	YES	YES	SVMLinear	0.636
16	YES	YES	YES	YES	YES	SVMLinear3	0.6485

5

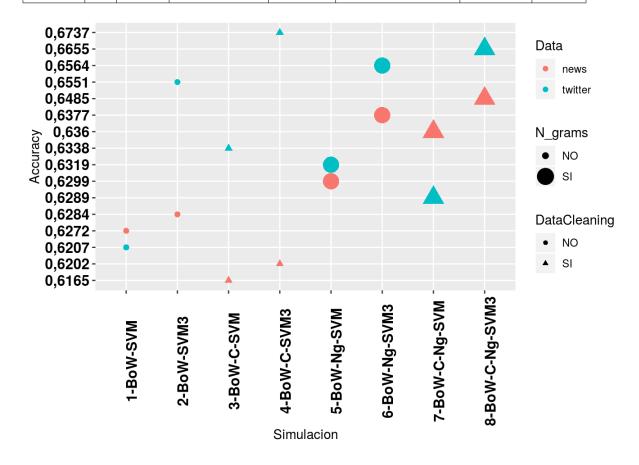


Fig. 2. Plot of the accuracy for every simulation done, considering the kind of text (Data), the use of N-grams (N-grams) and if the data have been cleaned (numbers, special characters and stop words). BoW=Back of Word, C=data cleaning (removing from texts numbers, special characters and Stop Words), Ng=N-grams, SVM=SVMLinear and SVM3=SVMLinear3.

Figure 2 shows graphically the results summarized on Tables 1 and 2. The labels of x axis means:

- **BoW-SVM/SVM3** → The model SVM Linear/SVM Linear 3 has been just trained with BoW.
- BoW-C-SVM/SVM3→ The model SVM Linear/SVM Linear 3 has been trained with BoW after removing from text numbers, special characters and Stop Words.
- BoW-Ng-SVM/SVM3→ The model SVM Linear/SVM Linear 3 has been just trained with BoW and N-grams without removing from texts numbers, special characters and Stop Words.
- BoW-C-Ng-SVM/SVM3→ The model SVM Linear/SVM Linear 3 has been just trained with BoW and N-grams removing from texts numbers, special characters and Stop Words.

In general we observe that SVM Linear 3 works better than SVM Linear. Also, in general, we observe that the accuracies are better for Twitter (less for case **BoW-C-Ng-SVM**) than for News Headlines. It could be due to News Headline text length are shorter than Twitter text length, this gives us a training corpus of words very different.

On the other hand, while we observe a certain growth trend in News Headlines with the methods used to train the model, for Twitter case this tendency is not so clear. Therefore, we should use different methodologies, adapting them to the characteristic of the text.

Finally, we observe, in a generalized way, that the accuracy is affected when Data Cleaning is applied. We think that this effect comes from the use of library called "**arabicStemR**" to remove Stop Words. Maybe, it could be solve using a set of Stop Words more adapted to Arabic language used in Twitter.

## Forward works

Considering the results obtained above, we have detected several deficiencies which have to be improved, such as:

- Study much deeper the effect of Data Cleaning: detect character more common in twitter written in Arabic and new set of Stop Word.
- Adjust much better the different parameter of SVM model in R-code.
- Study deeper the effect of N-gram.
- Adapt the methodologies to the characteristics of the text.

On the other hand, we propose some future steps to improve the model, such as:

- Detection and check the URLS.
  - This kind of structure is very common in Twitter, so if the twitter share an URL of a fake news o fake website we can consider the twitter as deceptive.

Detection of deceptions in Twitter and News Headlines written in Arabic

- Detection and checking the hashtags and Twitter accounts found on the texts, trying to see if them are related with fake hashtags and Twitter accounts.
  - This kind of structure are very common in Twitter, so if the twitter retweet a hashtag which is fake, we can consider the twitter as deceptive. The same for Twitter accounts.
- The study of N-grams different of 2-grams and 4-grams.
  - It would be interesting to study N-grams with other lengths.

## Conclusions

We have observed that the process of cleaning data can be affected negatively the accuracy of the model since we could remove relevant information. Mainly, News Headlines are affected stronger than Twitter. The explanation can be base on the number of words in Twitter is bigger.

Considering our approach, we have observe that just the use of N-gram improves the model, mainly in News Headlines, showing that there are structure of two and four word which can be indicator of deceptive text for this type of texts. In other words, this would be able to show that N-grams structures have a special characteristics in deceptive text for this type of texts.

On the other hand, we see that the methodologies have to be adapted to the characteristics of the text.

Finally we have proposed several improvements which can give a way to new jobs in the same line as this work, such as: detection and check of fake URLs, hashtag or Twitter accounts.

### References

- 1. Bernard Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read", https://www.forbes.com/
- 2. Jamie, "84 thoughts on "65+ Social Networking Sites You Need to Know About", https://makeawebsitehub.com/social-media-sites/
- 3. Brian Patrick Eha, "An Accelerated History of Internet Speed (Infographic)", https://www.entrepreneur.com/
- 4. Rajiv, "Evolution of wireless technologies 1G to 5G in mobile communication", https://www.rfpage.com/
- 5. Maria Valera, "Historia de Twitter: de un comienzo brillante a los rumores sobre su futuro incierto", https://marketing4ecommerce.net/
- 6. Lautaro Rubbi, "The Age of Disinformation", https://www.thebubble.com/
- 7. Rasha El Hallak, "In an era of information abundance, how much disinformation can we handle?", https://blog.usejournal.com/
- Meredith Wilson, "Disinformation is changing the way we view the world", https: //emergentriskinternational.com/

#### 8 F. E. Blázquez, M. Conde and J. M. Escalante

- 9. Jaume Masip, "Deception detection: State of the art and future prospects" Psicothema, Vol. 29, No. 2, pp. 149-159 (2017).
- P. Ekman, M. O'Sullivan, and M. Frank. A Few Can Catch A Lair. In Psychological Science, pages 10:263–266 (1999).
- M.L. Newman, J.W. Pennebaker, D.S. Berry, and J.M. Richards. Lying Words: Predicting Deception from Linguistic Styles. In Personality and Social Psychology Bulletin, pages 29:665–675 (2003).
- D.P. Twitchell, Jr J.F. Nunamaker, and J.K. Burgoon. Using Speech Act Profiling for Deception Detection. In Intelligence and Security Informatics: Second Symposium on Intelligence and Security Informatics, ISI 2004, pages 403–410 (2004).
- B.M. DePaulo, J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton, and H.Cooper. Cues to Deception. In Psychology Bulletin, pages 9:74–118 (2002).
- P. Ekman. Why Lies Fail and What Behaviors Betray A Lie. In J.C. Yuille (Ed.) Credibility Assessment, pages 71–81 (1989).
- L. Zhou, D.P. Twitchell, T. Qin, J.K. Burgoon, and Jr J.F. Nunamaker. An Exploratory Study into Deception Detection in Text-Based Computer Mediated Communication. In Proceedings of the 36th Hawaii International Conference on Systems Science (2003).
- 16. D.P. Biros, J. Sakamoto, J.F. George, M. Adkins, J. Kruse, J.K. Burgoon, and Jr. J.F. Nunamaker. A quasi-experiment to determine the impact of a computer based deception detection training system: The use of Agent 99 trainer in the US military. In Proceedings of the 38th Hawaii International Conference on Systems Science, volume 1 (2005).
- L. Zhou, J.K. Burgoon, Jr J.F. Nunamaker, and D.P. Twitchell. Automating Linguistic-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communication. In Group Decision and Negotiation Vol. 13, pp:81–106 (2004).
- J.W. Pennebaker, M.E. Francis, and R.J. Booth. Linguistic Inquiry and Word Count (LIWC). Technical report, Erlbaum Publishers, 2001.
- Rangel, F., Rosso, P., Charfi, A., Zaghouani, W., Ghanem, B., Snchez-Junquera, J.: Overview of the track on author profiling and deception detection in arabic. In: Mehta P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. In: CEUR-WS.org, Kolkata, India, December 12-15 (2019).
- 20. V. Vapnic, "The Nature of Statistical Learning Theory", Edt. Springer (1995).
- C. Cortez and V. Vapnic, "Support Vector Networks", Machine Learning Vol. 20, pp:273-297 (1995).
- Yang Y. M. and Liu X., "A Re-Examination of Text Categorization Methods", Proceeding of SIGIR-99,22 nd ACM International Conference on Research and Development in Information Retrieval (1999).
- Hu Zhang, Shan-de WEei, Hong-ye Tan and Jia-heng Zheng, "A Study on Deception Detection Based on Classification for Chinese Text", Journal of Computational Information Systems Vol. 3 No. 5, pp: 1-6 (2007)
- A.nand Rajaraman and J. Ullman, "Mining of Massive Datasets", Ed. Cambridge 2nd Edition (2010)
- Documentation of "arabicStemR", https://www.rdocumentation.org/packages/ arabicStemR/versions/1.2
- 26. Rahul Saxena, "Support Vector Machine Classifier Implementation in R with caret package", https://dataaspirant.com/