

UPV-INAOE-Autoritas - Check That: Preliminary Approach for Checking Worthiness of Claims

Bilal Ghanem¹, Manuel Montes-y-Gómez³,
Francisco Rangel^{1,2}, and Paolo Rosso¹

¹ PRHLT Research Center, Universitat Politècnica de València
{bigha@doctor, proso@dsic, fraranpa@prhlt}.upv.es

² Autoritas Consulting, Valencia, Spain

³ Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico
mmontesg@inaoep.mx

Abstract. Journalists usually work for a long time to investigate presidential debates. Their main role is to extract the sentences in the debates that include information about facts or previous events. These sentences are called claims. The investigation process of these claims is important where it can reveal how credible is the speaker or the other candidates. Therefore, proposing systems for extracting these claims can certainly improve the press work. In this paper, we will present our approach for CLEF-2018 Check That lab for Task 1. We propose an approach that uses a text distortion technique to detect claims that are worthy for checking. Our approach has achieved an acceptable results taking into account the complexity of this task.

Keywords: Factual Claims, English, Arabic.

1 Introduction

In the recent years, the political situation in many countries started to be more complex, which made the politicians started to exchange accusations in the public political debates, especially in the presidential debates. The prevailing situation in any debate is that each presidential candidate has a short period to respond to the claims of the other candidates. Each of them has the right to accuse the other with different claims, where the main objective is to convince the audience about his ability for that political position. During these long debates, the journalists' role is to investigate and validate the mutual accusations between those candidates to reveal the truth behind each claim. This task is complex to be done manually and in a short time, since the debates are very long, and each candidate declares many claims. Many of these claims are just opinions while the others contain previous facts that happened before. Starting from this issue, recent research topics are moving towards detecting check worthy claims in the presidential debates, where automatic approaches will save a lot of

time to journalists for this process. In this paper, we present our approach for detecting claims that are worthy for checking, we tried to predict factual claims by highlighting specific cue words in a classification process. In [1] we presented our approach for task 2 (checking claims factuality). In the following section, we present the previous works in the literature, then in Section 3, we describe the task. In Section 4, we describe our approach. Section 5 presents the experimental results, and finally in Section 6 we draw some conclusions and discuss further work.

2 Related Work

The main orientation in the literature was focused more on another related issue on claims, about detecting their factuality. Some researchers have proposed multiple approaches using different types of features to handle this issue. In [2] the authors have used a tree kernel approach to detect claims in the argument mining domain. Their approach used consistency trees of sentences to detect claims by capturing the trees similarities. They employed a support vector machine (SVM). In a different direction, a set of textual features to detect worthy claims in the presidential debates was proposed in [3]. They used sentiment polarity, number of words in a sentence, Bag-of-Words (BOW) using the Tf-Idf weighting scheme, Part-of-Speech tags, and named entities types. Since they generated a large feature set, they used a feature selection technique to extract the top N important features. Finally, different classifiers were tested, where the Random Forest classifier shows better results. Another related approach was proposed in [4], similarly, also for the presidential debates. The authors proposed different types of features, such as sentence-level features (sentiment, named entities, linguistic features etc.), contextual features (the position of a claim in the debate), and other mixed features such as text embedding, the discussed topic, and contradictions in the debate. For their approach, a SVM and a deep forward neural network (FNN) were used. In the results, the FNN showed better results comparing to the SVM. They created a corpus from the USA presidential debates of 2016, with a total of 5,415 sentences and they made it available to the research community.

3 Task Description

A set of presidential debates from the US presidential election are presented for the task, where each claim in the debate text has been tagged manually as worth to be check (1) or not (0). The text of the debates is used for the task as it is, to give the opportunity to the participants to exploit also contextual features in the debates. These debates are provided in two languages, English and Arabic, where the Arabic text is obtained translating from the English debates. The dataset that was provided is totally imbalanced, where the total number of claims is 4064: 90 claims are worth to be check and 3970 are not. The task goal is to detect the claims that are worthy for checking and to rank them, from the most

worthy one for checking to the lowest one. The Average Precision was used as a performance measure. More details of the task are mentioned in [8].

4 Proposed Approach

Previously, the authors in [5] have proposed a text distortion technique to enhance thematic text clustering by maintaining the words that have a low frequency in a document. Later on, a similar research in [6] has used the same text distortion technique for authorship attribution task, where the author has maintained the words that have the highest frequency in the documents, in an attempt to detect the author from his writing style.

We believe that this type of tasks is more thematic than stylistic, where the writing style is not important as the thematic words. In our approach, we used the same text distortion technique to detect worthy claims, where we concealed words that have high frequency in documents and maintaining (highlighting) other cue words that are used more in factual claims. Therefore, we followed [5] in their approach by concealing the most frequent words.

4.1 Text Distortion with Linguistic Features

In our approach, we have maintained the thematic words (that have the lowest frequency) using a threshold (C). The higher value of C is, the more thematic words are maintained. Also, we maintained a set of linguistic cue words (LC) that were used previously in [7] to infer the credibility of news (see Table 1). Additionally, we maintained also the named entities (NE) from being distorted, such as: Iraq, Trump, America. Through the manually checking of the claims, we found the checking worthy claims tended to list different types of named entities.

Table 1. Samples from the linguistic lexicons.

Linguistic lexicons	Examples
Assertives	appear, declare, guarantee, hypothesize
Factives	learn, realize, know, discover
Hedges	almost, guess, indicate, mostly
Implicatives	cause, manage, hesitate, neglect
Report	admit, answer, clarify, comment
Bias	adhere, act, agree, allow, addition
Subjectivity	afraid, champ, apologist, amusement

In Table 2, we show an example of the distortion process.

After applying the distortion process, the new version of the text was used by the char n-gram model using Tf-Idf weighting scheme. The new distorted text, become less biased by the high frequency words, such as stopwords. Finally, after preparing the distorted text, there is still one issue which is the value of C variable. The value of C is crucial, being a threshold between the amount

Table 2. An example of the text distortion process using different values of C.

Original claim	It was actually \$1.7 billion in cash, obviously, I guess for the hostages.
C = 0	** *** ***** \$ # . # ***** ** **** , ***** , * ***** ** ** ***** .
C = 0 & LC+NE	** *** actually \$ # . # ***** ** **** , obviously , * guess *** ** ***** .
C = 2000 & LC+NE	** *** actually \$ 1 . 7 ***** ** cash , obviously , * guess *** ** hostages .
C = 3500 & LC+NE	It *** actually \$ 1 . 7 billion in cash , obviously , I guess for *** hostages .

of thematic and the stylistic words. In Section 5, we show how we select the most appropriate value of C. For the Arabic language, we employed the same approach, where the only issue we had was the Arabic version of the linguistic lexicons. The manual translation of them is a time-consuming process, where they are quite large. Therefore, we used Google Translation API to translate these lexicons.

5 Experiments and Results

In this section, we present the tuning process of our approach. We carried out many experiments to test different machine learning classifiers. We found that the K-Nearest-Neighbor (KNN) has achieved the highest Average Precision value. We had two parameters to select the best model: the value of K-neighbors (K) of the KNN classifier and the C value of the distortion ratio. The selection process of these two values is hard to be set manually, therefore, we used the Grid Search technique to select the most appropriate values of these two parameters. The best value of K is 1, where for C value is 1700. The low value of K is due to the highly imbalanced situation of the dataset; larger values tend to bias the classifier to the majority class. A similar process was applied to select the best parameters but using word n-gram rather than character. For the evaluation, the Average Precision @N was used. The results of both runs are showed in Table 3. From these runs, we can see that char n-gram model outperformed clearly the one using word n-gram.

Table 3. The results obtained during the tuning phase using word and char n-gram models. We chose @N in our experiments as the last record in the testing part.

Approach	Classifier	K	C value	n-gram	AVG Precision @N
Text Distortion + char n-gram	KNN	1	1700	4	0.234
Text Distortion + word n-gram	KNN	2	2500	1	0.157
Baseline, word n-gram	SVM	-	-	1	0.163

Ranking Claims After the claims have been detected, it is important to rank

them based on their worthiness for checking. For the ranking process, we used the KNN classifier. We ranked the claims based on the KNN confidence in the classification process. At the beginning, we extracted the distances to the nearest neighbor (since we used K-neighbor equal to 1) for all the predictions in the test file. Then we applied a normalization for the distances to range 0-1. For each predicted instance, we checked the class type of the nearest neighbor: if it was positive, we subtracted the distance value from 1 and we used it for the ranking. We subtracted the distance from 1 to take the inverse of it: the small distance value (near to zero) mean a high classification confidence. The highest value (near to 1) is the one that obtained a higher rank (more worthy for checking). We applied the same process when the nearest neighbor is from the negative class, the rank value by -1, in order to discriminate the positive and the negative instances.

As we mentioned before, the used measure for this task is the Average Precision. In the official testing phase, multiple testing files were presented. For the final results the Mean Average Precision (MAP) was used. The official results of the task 1 are shown in Table 4. In the English part of the task, our approach

Table 4. Official results for the Task1, released using MAP measure.

Team	English	Arabic
Prise de Fer	0.1332	-
Copenhagen	0.1152	-
UPV-INAOE	0.1130	0.0585
bigIR	0.1120	0.0899
Fragarach	0.0812	-
blue	0.0801	-
RNCC	0.0632	-

has achieved the third position among seven teams, where the results are close to each other. In the Arabic part, only two teams have submitted their results. Similar to English, the results are close and there is not big difference between them. We believe that the lower results of our approach in the Arabic part is because of the automatic translation of the lexicons. A manual translation would have been more reliable.

6 Conclusion and Future Work

Detecting the claims that are worthy to be checked is important being a preliminary step for detecting their factuality. With these two steps, we can improve the journalists work of manual investigating for instance presidential debates. As a result of that, journalists can achieve their work quicker and in an easier manner. As we saw from the official results, performances are low, showing the difficulty of the task. Also, we can conclude that text distortion method worked

better than using the full text in the classification process, where it has improved the results comparing to the baseline with the normal BOW method. As future work, we will try to test more features that could discriminate the claims that are worthy to be checked from those that are not.

7 Acknowledgements

The authors acknowledge the SomEMBED TIN2015-71147-C2-1-P MINECO research project. The work on the data in Arabic as well as this publication were made possible by NPRP grant #9-175-1-033 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the last two authors.

References

1. Ghanem, Bilal, Manuel Montes-y-Gòmez, Francisco Rangel and Paolo Rosso. UPV-INAOE-Autoritas - Check That: An Approach based on External Sources to Detect Claims Credibility. In Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, CLEF '18, Avignon, France, September.
2. Lippi, Marco, and Paolo Torrioni. Context-Independent Claim Detection for Argument Mining. In IJCAI, vol. 15, pp. 185-191. 2015.
3. Hassan, Naeemul, Chengkai Li, and Mark Tremayne. Detecting Check-Worthy Factual Claims in Presidential Debates. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1835-1838. ACM, 2015.
4. Gencheva, Pepa, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, and Ivan Koychev. A Context-Aware Approach for Detecting Worth-Checking Claims in Political Debates. In Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, pp. 267-276. 2017.
5. Granados, Ana, Manuel Cebrian, David Camacho, and Francisco de Borja Rodriguez. Reducing the Loss of Information Through Annealing Text Distortion. IEEE Transactions on Knowledge and Data Engineering 23, no. 7 (2011): 1090-1102.
6. Stamatatos, Efstathios. Authorship Attribution using Text Distortion. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, vol. 1, pp. 1138-1149. 2017.
7. Mukherjee, Subhabrata, and Gerhard Weikum. Leveraging Joint Interactions for Credibility Analysis in News Communities. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 353-362. ACM, 2015.
8. Atanasova, Pepa and Màrquez, Lluís and Barrón-Cedeño, Alberto and Elsayed, Tamer and Suwaileh, Reem and Zaghoulani, Wajdi and Kyuchukov, Spas and Da San Martino, Giovanni and Nakov, Preslav, Overview of the CLEF-2018 Check-That! Lab on Automatic Identification and Verification of Political Claims. Task 1: Check-Worthiness. In Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, CLEF '18, Avignon, France, September.