

# CheckThat! Automatic Identification and Verification of Claims: IIT(ISM) @CLEF'19 Check Worthiness

Ritesh Kumar, Shivansh Prakash, Shashank Kumar and Rajendra Pamula

Department of Computer Science and Engineering,  
Indian institute of Technology (Indian School of Mines) Dhanbad, 826004  
India  
{ritesh4rmrvs,helloshivanshprakash@gmail.com,shashank0218@gmail.com,  
rajendrapamula}@gmail.com

**Abstract.** This paper describes the work that we did at Indian Institute of Technology (ISM) towards CheckThat!: Automatic Identification and Verification of Claims for CLEF 2019. As per requirement of CLEF-2019 we submit the only one run in its Check Worthiness Task. Two-fold cross-validation is used to select a model for submission to CheckThat Lab at CLEF 2019. For our run, we use SVM and LSTM method. Overall, our performance is not satisfactory. However, as new entrant to the field, our scores are encouraging enough to work for better results in future.

**Keywords:** LSTM, SVM, Feature extraction

## 1 Introduction

Investigative journalists and volunteers have been working hard to get to the core of a claim and present solid evidence in favor or against it. And in this day and age of information abundant amount of info/data is readily available, thus manual fact-checking is very time-consuming and therefore automatic methods were proposed as a means of speeding up the process [1]. Also, some steps of the fact-checking pipeline are given less attention, e.g. checking worthiness estimates is severely mis-understood as a problem. That is why we took this problem of checking worthiness estimates of statements/claims. The overview of of shared task for CheckThat! can be found in [2]. In brief, we have to identify which sentence should be prioritized for fact-checking given a political debate or a transcribed speech, segmented into sentences with annotated speakers. This is a ranking task and systems are required to produce a score per sentence that will perform the ranking. The task is performed in English.

The rate at which a statement in an interview, a press release, or a tweet can spread almost instantly across the globe has created an unprecedented situation

---

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2019, 9-12 September 2019, Lugano, Switzerland.

[3,4]. There is almost no time to cross-check a statement or a claim against facts and this has proved critical in politics, e.g. during the 2016 US Presidential Campaign, whose results is said to be affected by fake news and claims spread using social media. As it became apparent that this is a problem that can create great affects in our lives, a number of fact-checking programs have started, led by organizations such as FactCheck and Snopes [5–7].

The organization of the rest of the paper is as follows. Section 2 describes about the dataset. We describe our methodology: field categories and indexing, which document and topic fields we used for retrieval in Section 3. In Section 4 we describe our results. Finally, we conclude in Section 5 with directions for future work.

## 2 Data

We participate in Task 1. The detail description of dataset can be found in [8]. The training data consisted of 19 files which include political debates and speeches. Each file contains the debate/speech split into sentences. Each line contains a single sentence, its speaker and a label which is annotated by experts as check-worthy or not. The sentence is labelled 0 if it is not worthy of checking and 1 if it is worthy for checking. The data consists of a total of 16,421 sentences, of which 440 were labeled as check-worthy. There is an imbalance in the dataset with only 2.68% of the dataset labelled same as that of the target class i.e. check-worthy. A few instances of this training data, along with their speakers and labels, are presented below :

300 SANDERS Let’s talk about climate change. 0

301 SANDERS Do you think there’s a reason why not one Republican has the guts to recognize that climate change is real, and that we need to transform our energy system? 1

The test data is a collection of seven files consisting of debates and speeches. In this task, we do not use any external knowledge other than domain-independent language resources such as parsers and lexicons. Instead, we concentrate on extracting linguistic features that can indicate the check-worthiness of the sentences .

## 3 Methodology

### 3.1 Data Preprocessing and Feature Extraction

Data Processing is required to prepare data to be used by systems to classify and rank statements according to the check-worthiness score. Rich feature extraction is needed to perform ranking according to language constructs rather than relying on heuristic or encyclopedia Knowledge. Syntactic and semantic features were extracted from both speeches and debates to represent sentences consistently and converted every sentence into vector. These features are: Lexical Features, Sentence Embedding, Stylometric Features, Semantic Features,

ffective Features and Metadata Features. First of all we perform speaker normalization by assigning each speaker an unique Id. The reason for this is that the same speaker is present in the training data with different names in different instances. The sentences in the training data are first tokenized. We also remove stop-words and the remaining tokens are stemmed using a stemmer. A single file which contains sentences extracted from required file is made to be used as training data. Syntactic and semantic features are extracted from both speeches and debates to represent sentences consistently and converted every sentence into vector.

We use two approaches for designing the model for this task Support Vector Machine (SVM) and Recurrent Neural Network using Long Short Term Layers (LSTM). We use SVM because of its simplicity, ease of use and its ability to avoid over-fitting using regularization parameters. However, since the datasets are more aptly described using a sequence of sentences, a model which considers this sequence-time nature of the data will obviously be a better choice. Since SVMs do not consider this, we also try to implement a recurrent neural network model having memory units, in the form OF LSTM layers.

For our model, we use a layer of LSTM cells having an input shape of (batch size, no. of time steps, feature dimension). There were 64 units of LSTM cells in each layer hence, the output shape of the layer will be (64). We then add a dense layer of 64 neurons after the LSTM layer with a Rectified Linear Unit (ReLU) activation function. A dropout layer is also added to avoid over-fitting of data. Finally, a softmax output layer is added producing output probabilities of the classes for each input instance.

## 4 Results

The complete training data is divided into two parts i.e. training data and validation data. The validation data consists of a speech and a debate selected from the training part. The models are trained using the training part and evaluated using validation part. Many models with different parameters are generated and evaluated and the model giving the best results is selected from them. We are provided with some code that helped us to evaluate our models in various evaluation metrics. We get the best result using Recurrent neural network model rather than SVMs. The scores obtained by our six runs are given in Table 1. The official evaluation measure provided by CLEF'19 is *MAP*. We show the best score in the task demonstrated by run-id **Copenhagen(\*)**, for the sake of comparison.

**Table 1.** Results - The official evaluation Measure by CLEF 2019

<b>RUN ID</b>	Rank	<b>MAP</b>	<i>RR</i>	<i>R - P</i>
<b>Copenhagen(*)</b>	1	<b>.1660</b>	.4176	.1387
<b>ISMD16titlefield</b>	10	.0835	.2238	.0714

## 5 Conclusion and Future Work

This year we participate in Task 1 of the Check That!: Automatic Identification and Verification of Claim. A SVM and a Recurrent Neural Network model is used with supervised learning to classify and rank sentences according to their check-worthy score in political debates and speeches. A rich feature set is extracted to represent sentences in best possible way to tackle class imbalance and avoid heuristic or encyclopedia Knowledge. Two-fold cross-validation is used to select a model for submission to CheckThat Lab at CLEF 2019. While there can be no denial of the fact that our overall performance is average, initial results are suggestive as to what should be done next. Our work has shown us lot of interesting possibilities for future work. There is a lot of room for improvements in our task. Linguistic form of information is under-studied and can be explored in more depth for better result. We use shallow syntactic features in our project, this can be improved by using very deep syntactic feature to represent sentence. Further study is required for including more linguistic and non-linguistics features which may affect check-worthy score of phrases. Furthermore, we can use more complex recurrent neural network models to achieve better result. We shall be exploring some of these tasks in the coming days.

## References

1. Recasens, M., Danescu-Niculescu-Mizil, C., Jurafsky, D.: Linguistic Models for Analyzing and Detecting Biased Language. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 1650-1659 (2013).
2. Elsayed, Tamer and Nakov, Preslav and Barrón-Cedeño, Alberto and Hasanain, Maram and Suwaileh, Reem and Da San Martino, Giovanni and Atanasova, Pepa. Overview of the CLEF-2019 CheckThat!: Automatic Identification and Verification of Claims. Experimental IR Meets Multilinguality, Multimodality, and Interaction", LNCS, Springer, Lugano, Switzerland, September, 2019.
3. Porter, M.F.: Snowball: A Language for Stemming Algorithms. <http://snowball.tartarus.org/texts/introduction.html> (2001)
4. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. pp. 7986. Association for Computational Linguistics (2002)
5. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781 (2013)
6. Loria, S.: TextBlob: Simplified Text Processing. <http://textblob.readthedocs.org/en/dev/> (2014)
7. Cazalens, S., Lamarre, P., Leblay, J., Manolescu, I., Tannier, X.: A content management perspective on fact-checking. In: " Journalism, Misinformation and Fact Checking" alternate paper track of" The Web Conference" (2018)
8. Atanasova, Pepa and Nakov, Preslav and Karadzhov, Georgi and Mohtarami, Mitra and Da San Martino, Giovanni. Overview of the CLEF-2019 CheckThat! Lab on Automatic Identification and Verification of Claims. Task 1: Check-Worthiness, CLEF-ceur, 2019