

Self-Supervised Claim Identification for Automated Fact Checking

Archita Pathak **Mohammad Abuzar Shaikh** **Rohini K. Srihari**
University at Buffalo (SUNY) University at Buffalo (SUNY) University at Buffalo (SUNY)
Buffalo, NY Buffalo, NY Buffalo, NY
architap@buffalo.edu mshaikh2@buffalo.edu rohini@buffalo.edu

Abstract

We propose a novel, attention-based self-supervised approach to identify “claim-worthy” sentences in a fake news article, an important first step in automated fact-checking. We leverage *aboutness* of headline and content using attention mechanism for this task. The identified claims can be used for downstream task of claim verification for which we are releasing a benchmark dataset of manually selected compelling articles with veracity labels and associated evidence. This work goes beyond stylistic analysis to identifying content that influences reader belief. Experiments with three datasets show the strength of our model¹.

1 Introduction

The explosion of fake news on social media has resulted in global unrest and has been a major concern for governments and societies worldwide². According to a recent Pew Research Study, Americans rate it as a larger problem than racism, climate change, or illegal immigration³. Since, it’s inexpensive to create a website and easily disseminate content on the social media platforms, there is a rising need for automated fake news detection. Furthermore, AI solutions are also required to follow good practices, specifically avoiding censorship, violation of fundamental rights such as freedom of expression, and ensuring data privacy (de Cock Buning, 2018). However, to date, AI models proposed for fake news detection do not

scale for detecting real-time fake news⁴.

Much of the research on automated text-based fake news detection can be classified into three broad categories: (1) linguistic approach, which focuses on lexical, stylistic and pattern learning mechanisms (Potthast et al., 2017; Rashkin et al., 2017; Wang, 2017; Singhanian et al., 2017; Pérez-Rosas et al., 2018); (2) network-based approach, which leverages features such as the speed and volume of propagation of fake news articles on social media platforms (Castillo et al., 2011; Yang et al., 2012; Kwon et al., 2013; Ma et al., 2015; Jin et al., 2016; Ruchansky et al., 2017; Wu and Liu, 2018); and (3) automated fact-checking approach, which is an effort to assist manual fact-checkers by automating some of their tasks such as detection and verification of claims (Graves, 2018).

While most work in automated fact-checking has been focused on claim verification task, very few methods have been proposed for detection of claims (Hassan et al., 2017; Jaradat et al., 2018; Konstantinovskiy et al., 2018). The approaches in these efforts are majorly related to *political discourse*. However, our focus is on *fake news*, which are broader than *political discourse* since (i) they are deliberately written with a divisive agenda to cause social unrest, (ii) they are not constrained to only politics, and (iii) the headline plays an equally important role in compelling people to read the article.

In this paper, we focus on articles where there is a deliberate intent to influence readers through fabricated or manipulated claims in the headline and the content. Such articles have a compelling writing style similar to the mainstream media. Hence, we build datasets containing these type of compelling articles along with veracity labels and associated evidence supporting the label of each arti-

¹Data and code available at: <https://github.com/architapathak/Self-Supervised-ClaimIdentification>

²<https://www.reuters.com/article/us-singapore-politics-fakenews-factbox/factbox-fake-news-laws-around-the-world-idUSKCN1RE0XN>

³<https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/>

⁴<https://www.technologyreview.com/s/612236/even-the-best-ai-for-spotting-fake-news-is-still-terrible/>

cle. We, then, use these datasets to identify “claim-worthy” sentences. In our work, we define “claim” as “*statements which are important to the point of the article but one would require to have them verified.*”

Our working hypothesis is that in fake news which are created to cause harm, these are the sentences most relevant to the headline. Exploiting the hypothesis that the essence of a news article is encapsulated in its headline (Jaime Sis and MERCEDES, 2009; Kuiken et al., 2017; Wahl-Jorgensen and Hanitzsch, 2009), we propose a self-supervised method to explore the *aboutness* of the content with the headline of the article to extract the most relevant sentences. Bruza and Huibers (1996) defines *aboutness* as: *an information carrier i will be said to be **about** information carrier j if the information borne by j holds in i.* The idea is taken from Information Retrieval domain where it is used to signify implications between query and document, specifically to explore the underlying meaning or concept within the document and the query (Azopardi et al., 2009). In our work, headline is modelled as a query while each of the sentences of the article acts as a document, and we use the concept of *aboutness* to find the relevant sentences. We show that attention-based mechanisms are able to successfully capture this concept in the news article.

Contribution: In this work: (i) we introduce a self-supervised representation learning model that eliminates the prerequisite that requires human to annotate data, which is a time consuming and costly task; (ii) the proposed headline-to-sentence attention-based approach for claim identification is novel; previous unsupervised approach for this task use weak supervisory signal which does not capture the context of the article efficiently; and (iii) we propose a benchmark dataset for evidence-based fake news detection. Our dataset contains evidence for each of the fake news articles that contributes to the overall degree of veracity of the article.

2 Related Work

Claim Identification/Detection: The task of claim identification/detection was first introduced by Levy et al. (2014) who, with the help of human annotators, provided a dataset and a fundamental approach in identifying context dependent claims. In their dataset, which was originally developed by Aharoni et al. (2014), each statement indicates

whether it should be considered as a context dependent claim (CDC) or not. Levy et al. (2014) reported encouraging results obtained through a supervised learning algorithm using a cascade of classifiers. A rule-based model was introduced by Eckle-Kohler et al. (2015) to bifurcate claim and premise statements in an argumentative discourse environment. However, these methods were generic to only a small set of corpora. Furthermore, Levy et al. (2017) also introduced an unsupervised approach to detect claims, which involves a weak supervisory signal “that” for training. However, this approach does not capture the *aboutness* of the article to understand the context of “claim-worthy” sentences.

In 2017, Hassan et al. (2017) introduced Claim-Buster, a platform developed by training a supervised learning model on a large annotated corpus of televised debates in the USA. Their model used SVM classifier to detect *claim-worthy factual claims* and produced a score of how important a claim is to factcheck. The 20,000 sentences in the corpus were annotated by human coders to distinguish between claim-worthy factual claims from opinions and boring statements. However, annotating a sentence as an important or unimportant claim is a non-trivial task as this decision changes depending on who’s asking, political context and annotator’s background (Graves, 2018).

The model proposed by Hassan et al. (2017) only learns the labelled instances and does not explore the contextual information of the written text. A context-aware approach in the political discourse environment was introduced by Gencheva et al. (2017) who created a rich representation of the sentences from 2016 US presidential debates. Their dataset was compiled by taking the outputs of factchecking of the debates from 9 factchecking organizations. Their models were created to predict if the claim would be highlighted by at least one or by a specific organization. However, the authors don’t have any formal definition of claim in their paper, and their model is specific to certain organizations which led to several false positives.

Another context-aware approach for claim detection was proposed by Konstantinovskiy et al. (2018) who used sentence embeddings, pre-trained on a large dataset of NLI. This work also created a crowd-sourced annotated dataset of sentences from UK political TV shows, annotated across 7 classes. However, their classifiers for the fine-grained clas-

sification to detect 7 classes of sentences did not yield good results due to lack of enough annotated data, thus requiring more annotations which is a costly and time consuming task.

We build a model that can be trained in a self-supervised setting to overcome the challenges associated with annotated dataset of claims. We also use attention-based approach to capture *aboutness* and rich contextual information between headline and all the sentences of the article. The performance on manually created test sets demonstrate promising results in identifying “claim-worthy” sentences even when no sentence-level annotation was used for training.

Fake News Dataset: A variety of fake news datasets have been released in the recent years, most notably BuzzFeed⁵ and Stanford (Allcott and Gentzkow, 2017) datasets containing list of popular fake news articles from 2016 US presidential elections. However, these datasets only contain webpage URLs of the original article and majority of them don’t exist anymore. Following this, several other datasets were published such as Fake news challenge dataset⁶ which was used for the task of stance detection; *Getting Real about Fake News* Kaggle dataset⁷ which was created by using BS detector tool; and FakeNewsCorpus⁸ which is an open-source large scale collection of fake news articles. However, these articles are labelled as fake based on the domain of the websites they come from. Since, the content of these articles are not verified for degree of veracity, using them directly for training may lead to several false positives.

This problem was overcome by recently released large dataset, NELA-GT-2018 (Norregaard et al., 2019), which contains articles with ground truth ratings retrieved from 8 different assessment sites. However, the label definitions are not generic and dependent on the external organizations. Pathak and Srihari (2019) also introduced intuitive ground truth labels based on the degree of veracity of the fake news articles, however, the dataset is not publicly available. Additionally, they also do not specify the relationship of their labels with the labels used by established fact-checking organizations. Furthermore, due to lack of evidence in these datasets, they cannot be used for downstream task

⁵<https://www.buzzfeednews.com/article/craigsilverman/these-are-50-of-the-biggest-fake-news-hits-on-facebook-in>

⁶<http://www.fakenewschallenge.org/>

⁷<https://www.kaggle.com/mrisdal/fake-news>

⁸<https://github.com/several27/FakeNewsCorpus>

of evidence-based verification, which is one of the motivations of this paper. We overcome all these limitations in our datasets described in the following section.

3 Datasets

We introduce two datasets of compelling fake news articles which have writing style similar to mainstream media. The first dataset, DNF-700, where DNF stands for *DisiNFormation*, contains articles on politics published within 4 months of 2016 US Presidential Elections from questionable sources (non-mainstream). To compile this dataset, we first extracted fake news articles from working web page URLs of Stanford dataset (Allcott and Gentzkow, 2017). However, majority of webpage URLs in this dataset are expired and we could extract only 26 fake news articles. Therefore, we then used “*Getting real about fake news*” Kaggle⁸ dataset to sample more articles on politics. Since, most of the articles in this dataset contain anomalies (eg: incomplete article, social media comments labelled as fake etc.), we manually verified the writing style and discarded obvious fakes - articles with poor grammar and excessive usage of punctuations. However, the degree of veracity of each article in this dataset is not checked and some articles may contain personal opinions.

The second dataset, DNF-300, is more sophisticated subset of DNF-700, containing 290 compelling articles on Politics and 10 on Health/Medical news. Unlike other fake news datasets in which veracity and evidence for articles are not provided, DNF-300 contains articles associated with veracity labels as well as corresponding evidence. The process of annotating this dataset involves identifying sentences from each article based on their persuasive tone and relevance with the headline. These sentences were then queried on the web and top 10 results were considered to gather evidence from credible sources⁹. Based on the evidence found, we label the entire article into four categories: $\{(0) \text{ false}; (1) \text{ partial truth}; (2) \text{ opinions stated as fact}; (3) \text{ true}\}$. These labels are inspired by (Pathak and Srihari, 2019); Table-1 shows the description and distribution of these labels while the comparison with two popular fact-checking websites is displayed in Figure-1. An

⁹Credible sources were extracted from <https://mediabiasfactcheck.com/> The sources range between left, center and right biased news sources

Label	False	Partial True	Opinion Stated As Fact	True
Desc.	(i) No evidence could be found, or (ii) found evidence to refute the entire article	Article about true event, however, found evidence refuting some of the claims	Article contain false/manipulated claims, however, it's an opinion article which cannot be labelled as fake	Found evidence supporting the entire article
Total	126	75	79	20

Table 1: DNF-300 label description and distribution. Claims here are the sentences manually selected based on their persuasive tone and relevance with the headline. Interestingly, some of the articles, which were labelled as fake in other datasets due to the domain of publishing website, turned out to be true news.

example from the dataset is shown in Table-2

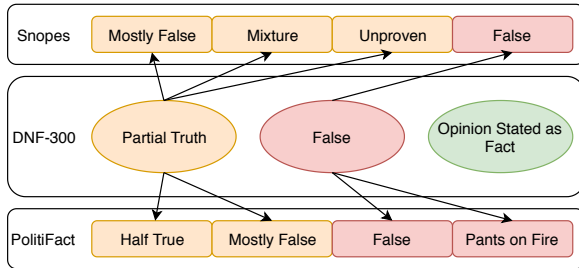


Figure 1: Label Comparison with Snopes and PolitiFact ratings.

This dataset is also a key contribution of this paper as the articles are manually read and verified. Additionally, the dataset contains two novel features which are essential for the fake news verification task: (i) generic veracity-based label set, independent of any external organization, and (ii) ground truth evidence corresponding to each label.

In addition to these two datasets, we also train our model for claim identification on the dataset introduced for context dependent claim detection (CDCD) by Levy et al. (2014). Although this dataset (CDC) does not contain fake news articles, it has manually annotated sentences based on their relevance to a certain topic. These annotations were utilized for the evaluation of our self-supervised learning model described in the following section. More details on the datasets and examples can be found in Appendix.

4 Architecture

Problem Definition: Given an article with a set of sentences $S = \{S_1, S_2, \dots, S_i, \dots, S_n\}$ and a headline H , the task of our multihead attention claim

identification network (MA-CIN) is to extract the sentence most relevant to the headline. Our self-supervised model exploits the rich contextual information to extract the relevant sentences which are considered as “claim-worthy”.

Approach: For this task, we implement two types of attention: (i) self-attention on all sentence vectors so that each sentence S_i is aware of all other sentences in S ; (ii) cross-attention of headline vector on each sentence vector, so that all self-attended sentences are also aware of the headline’s context. We then generate headline based on the context-aware sentences, and compare it with the original headline in three different settings as listed below:

1. **Headline Vector (MA-CIN (HV)):** In this setting, the original headline vector acts as the supervisory signal for self-supervised learning. We minimize the mean squared error (MSE) between the generated and the original headline vectors for training.
2. **Headline One-Hot Word Vector (MA-CIN (OHV)):** In this setting, the words in the original headline act as the supervisory signal. We use LSTM (Hochreiter and Schmidhuber, 1997) to predict at most 50 words, from a vocabulary of 20,000 words, to generate a one-hot-vector for each word of the new headline. We then minimize the categorical cross-entropy error (CCE) at each time step corresponding to each word in original and new headlines for training.
3. **Combined HV & OHV (MA-CIN (Combined)):** In this setting, both original headline vector and the words act as supervisory signal. Therefore, we combine the two loss functions mentioned above to train the model.

For this, we build several layers in our architecture (see Figure-2), which are delineated as follows:

4.1 Sentence Embeddings

Each sentence S_i in S , and headline H are converted to a fixed length 300-dimensional vector, s_i and h , such that $s_i, h \in R^{1 \times D}$, where $D = 300$. For uniformity, we calculate the maximum number of sentences L that an article can contain in the respective corpus. Next, we zero pad the difference in the quantity of sentence vectors in each article such that every article can be represented as a vector $A \in R^{L \times D}$.

Headline: Allergens in Vaccines Are Causing Life-Threatening Food Allergies

It would probably surprise few people to hear that food allergies are increasingly common in U.S. children and around the world . According to one public health website , food allergies in children aged 0-17 in the U.S. increased by 50% from 1997 to 2011. Although food allergies are now so widespread as to have become almost normalized, it is important to realize that millions of American children and adults suffer from severe rapid-onset allergic reactions that can be life-threatening. Foods represent the most common cause of anaphylaxis among children and adolescents. The United Kingdom has witnessed a 700% increase in hospital admissions for anaphylaxis and a 500% increase in admissions for food allergy since 1990. The question that few are asking is why life-threatening food allergies have become so alarmingly pervasive. A 2015 open access case report by Vinu Arumugham in the Journal of Developing Drugs , entitled “ Evidence that Food Proteins in Vaccines Cause the Development of Food Allergies and Its Implications for Vaccine Policy ,” persuasively argues that allergens in vaccinesand specifically food proteinsmay be the elephant in the room. As Arumugham points out, scientists have known for over 100 years that injecting proteins into humans or animals causes immune system sensitization to those proteins. And, since the 1940s, researchers have confirmed that food proteins in vaccines can induce allergy in vaccine recipients. Arumugham is not the first to bring the vaccine-allergy link to the publics attention. Heather Fraser makes a powerful case for the role of vaccines in precipitating peanut allergies in her 2011 book, The Peanut Allergy Epidemic: Whats Causing It and How to Stop It.

Type: 1 (*Partial Truth*)

Authors: Admin - Orissa

URLs: galacticconnection.com

Evidence: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3890451/>

Reason: The key claim is written in such a way so that it misleads people in thinking all the food related allergies in US are caused by vaccines. Found evidence which says these type of allergies are rare.

Table 2: An example on *Partial Truth* type from DNF-300 dataset.

4.2 1D Convolution

To effectively capture local relevance, we leverage 1D-CNN (LeCun et al., 1998) to extract the features from the article vector A . For our experiments the kernel size for each convolution layer is $K \times D \times C$, where K is kernel-width and C is the number of filters. This means the network will process K sentences at a time. The size of K and C is a hyper-parameter and as per our experiments, we set $K = 4$ with an assumption that not more than 4 consecutive sentences will be relevant to each other.

4.3 Self Attention

Inspired by the attention implementation in (Zhang et al., 2018; Vaswani et al., 2017), to capture global relevance, the article features from the previous 1D-CNN layer are transformed into feature spaces q, k to calculate attention, where $q(x) = W_q x$ and

$$k(x) = W_k x.$$

$$\beta_{j,i} = \frac{\exp(z_{ij})}{\sum_{i=1}^N \exp(z_{ij})}, \text{ where } z_{ij} = q(x_i)^T k(x_j) \quad (1)$$

$\beta \in R^{N \times N}$ is the attention coefficient, which is the normalized relevance score between the sentence x_i and x_j . β is then matrix multiplied by v , where $v(x) = W_v x$, to obtain the context rich output $o_j \in R^{C \times 1}$.

$$o_j = \sum_{i=1}^N \beta_{j,i} v(x_i), \text{ where } o_j \in \{o_1, o_2, \dots, o_N\} \quad (2)$$

Finally, the output of the self-attention layer is $o \in R^{C \times N}$, which is computed as

$$o_j = g(o_j), \text{ where, } g(x) = W_g x \quad (3)$$

In the above equations, $x \in R^{C \times N}$ is obtained after applying 1D convolution on sentence vectors, $W_q \in R^{\bar{C} \times C}$, $W_k \in R^{\bar{C} \times C}$, $W_v \in R^{\bar{C} \times C}$, $W_g \in$

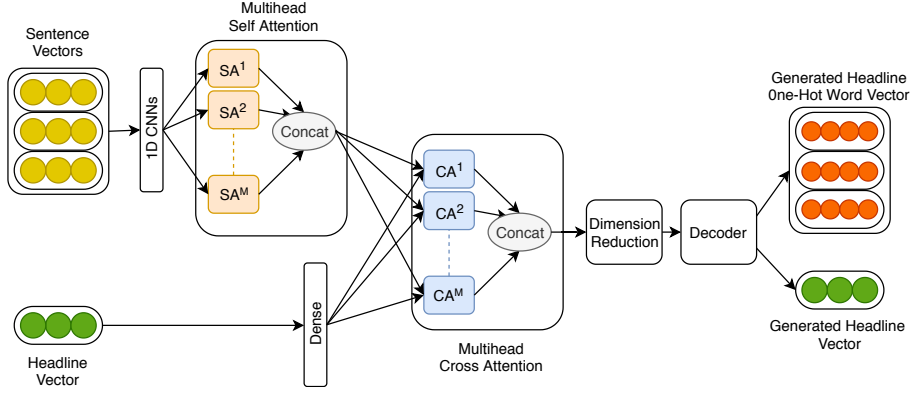


Figure 2: Architecture of Multihead Attention - Claim Identification Network (MA-CIN). The model is trained by using self-supervised learning approach using three variants of supervisory-signal - headline vector, headline words and the combination of both vector and words.

$R^{C \times \bar{C}}$ and output $o \in R^{C \times N}$. Following the work by (Zhang et al., 2018) we preferred the value of $\bar{C} = \frac{C}{8}$ for computation effectiveness. We also multiply a λ and γ , learnable scale parameters, to the output of our attention module and input vector respectively to allow the network to choose between local and global sentences effectively.

$$o = \gamma x + \lambda o \quad (4)$$

γ is initialized to 1 and λ is initialized to 0, so as to allow the local context to be captured effectively during the early iterations and as the value of λ increases it allows the network to add more context to the representation.

4.4 Multihead Concatenation

In the architecture, we could apply self attention to the input x M times resulting into M attention heads. The output of one attention head is denoted by o . We concatenate the outputs o^M to get a richer representation allowing the network to capture various relationships.

$$msa_o = \left\| \left\|_{i=1}^M o^i = o^1 \right\| \dots \left\| o^M \right\| \right\| \quad (5)$$

where, $msa_o \in R^{MC \times N}$ is the long range context aware output of multihead self attention. Here, $\|$ denotes concatenation across axis C .

4.5 Cross Attention

The headline vector is transformed into a feature space $\bar{h} = W_h h$, where $\bar{h} \in R^{\bar{C} \times 1}$ and then, it's relevance is calculated with msa_o , obtained from the previous layer, by using equations defined in

4.3. Finally, after applying multihead concatenation using 5, we obtain headline-context aware representation, $mca_o \in R^{MC \times N}$. We fix $M = 4$ for all our experiments.

4.6 Loss Function

To generate the headline vector d_h as close to the input headline vector h , we apply Mean Squared Error between d_h and h and calculate the headline vector generation loss L_v

$$L_v = \frac{1}{n} \left(\sum_{i=1}^n (d_{h_i} - h_i)^2 \right) \quad (6)$$

For estimating the probability of a word from the vocab in the predicted headline we calculate the cross-entropy between the predicted headline words d_{hw} and input headline one-hot vector HW .

$$L_w = - \sum_i d_{hw_i} \log(HW_i) \quad (7)$$

The total loss $L_{total} = L_v + L_w$ is then evaluated for all samples $b \in B$, where B is one batch.

5 Experiments and Evaluation

5.1 Training Setup

We train our Multihead Attention model for Claim Identification, MA-CIN, on datasets mentioned in Section 3. The CDC dataset contains total of 522 articles. Amongst these, there are 47 articles with 8 or more annotated claim sentences which are considered as evaluation set (CDC.Eval) for this dataset. Next, for DNF-300 and DNF-700, we asked two annotators to manually tag at least 5 sentences as ‘‘claim-worthy’’ in each of the 50 articles. Sentences which were consented by both

Dataset	Configuration	CDC_Eval			DNF_Eval		
		Prec.	Rec.	F1	Prec.	Rec.	F1
<i>Spacy</i>	Baseline	0.09	0.14	0.11	0.33	0.42	0.37
<i>CDC</i>	Baseline (Levy et al., 2014)	0.23	-	-	-	-	-
	MA-CIN(HV)	0.18	0.08	0.11	0.39	0.53	0.45
	MA-CIN(OHWV)	0.25	0.10	0.15	0.40	0.54	0.46
	MA-CIN(Combined)	0.26	0.11	0.16	0.42	0.57	0.48
<i>DNF - 700</i>	MA-CIN(HV)	0.20	0.09	0.12	0.37	0.54	0.44
	MA-CIN(OHWV)	0.19	0.08	0.11	0.40	0.5	0.44
	MA-CIN(Combined)	0.28	0.12	0.16	0.41	0.55	0.47
<i>DNF - 300</i>	MA-CIN(HV)	0.19	0.08	0.11	0.39	0.53	0.48
	MA-CIN(OHWV)	0.25	0.11	0.15	0.38	0.53	0.45
	MA-CIN(Combined)	0.24	0.10	0.14	0.42	0.57	0.48

Table 3: Comparison of MA-CIN model configurations over three datasets and two evaluation sets for identification of “claim-worthy” sentences.

Headline: Clinton Received Debate Questions Week Before Debate				
0	The first presidential debate was held and Hillary Clinton was proclaimed the winner by the media.	-	-	0.41
1	Indeed Clinton was able to turn in a strong debate performance, but did she do so fairly?	-	-	0
2	Multiple reports and leaked information from inside the Clinton camp claim that the Clinton campaign was given the entire set of debate questions an entire week before the actual debate.	GT	PD	1
3	Earlier last week an NBC intern was seen hand delivering a package to Clinton’s campaign headquarters, according to sources.	-	PD	0.73
4	The package was not given to secretarial staff, as would normally happen, but the intern was instead ushered into the personal office of Clinton campaign manager Robert Mook	-	PD	0.68
5	Members of the Clinton press corps from several media organizations were in attendance at the time, and a reporter from Fox News recognized the intern, but said he was initially confused because the NBC intern was dressed like a Fed Ex employee.	-	-	0.46
6	The reporter from Fox questioned campaign staff about the intern, but campaign staff at first claimed ignorance and then claimed that it was just a Fed Ex employee who had already left.	-	-	0.67
7	No reporters present who had seen the intern dressed as a Fed Ex employee go into Mook’s office saw him leave by the same front entrance.	-	-	0.49
8	The Fox reporter who recognized the intern also immediately looked outside of the campaign headquarters and noted that there were no Fed Ex vehicles parked outside.	-	-	0.51
9	Clinton seemed to have scripted responses ready for every question she was asked at the first debate.	GT	-	0.37
10	She had facts and numbers memorized for specific questions that it is very doubtful she would have had without being furnished the questions beforehand.	GT	-	0.63
11	The entire mainstream media has specifically been trying to portray Trump as a racist and a poor candidate.	-	-	0.24
12	By furnishing Clinton with the debate questions NBC certainly hoped to make Clinton appear much more knowledgeable and competent than Trump.	GT	PD	0.79
13	And though it is unlikely that anyone will be able to conclusively prove that Clinton was given the debate questions, it seems both logical and likely.	GT	PD	0.7

Figure 3: Interpretation of relevance of sentences with the headline of an example article from DNF-300. GT and PD indicate ground truth and top-5 predicted “claim-worthy” sentences, respectively. MA-CIN model was able to predict 3 most relevant sentences correctly. Last column shows the attention weights between headline and each of the sentences of the article. Sentence 2 has been correctly predicted as the most relevant while sentence 1 is the least relevant.

the annotators as “claim-worthy” were finalized as ground truth claims for these 50 articles, and used as testing set for evaluating the model performance on DNF datasets. The remaining 475 articles from CDC, 250 articles from DNF-300, and 650 articles from DNF-700 were split into 5 folds to train the model using a 5-Fold cross validation (Kohavi, 1995), where we use 4 folds for training and 1 fold for validation. Each of the three settings, described in Section- 4: MA-CIN(HV), MA-CIN(OHWV) and MA-CIN(Combined), was trained with each of the three datasets, and evaluated on DNF_Eval and CDC_Eval. Total number of parameters for these three settings are 15,012,916 (10,240 non-trainable), 40,975,656 (10,240 non-trainable) and 41,812,564 (12,288 non-trainable) respectively. All other network parameters are displayed in supple-

mental material.

In each setting, we use batch normalization, ReLU non-linearity as an activation function, and a dropout of 0.5 for every convolution operation. We trained all the models for 2000 epochs, where, for every training we used Adam optimizer with a learning rate $lr = 0.0001$, $\beta_1 = 0.99$ and $\beta_2 = 0.0$. There was no weight decay set as the model was trained in a self-supervised setting with finite epochs and an already small learning rate. Glove 300D word embedding was used for all our experiments and the number of input sentences was set to 500. The models were trained on three 11GiB Nvidia 1080Ti GPUs in parallel.

5.2 Evaluation Metrics

We evaluate MA-CIN models on two evaluation sets, DNF_Eval and CDC_Eval. With self-

supervised setting we first rank the sentences based on relevance with the headline and then extract the top five sentences along with their sentence ids as “claim-worthy” sentences. For evaluation on DNF_Eval, we calculate the *true positives*(TP), *false positives*(FP) and *false negatives*(FN) from ground truth claim ids. To evaluate on CDC_Eval, since we do not have ground truth claim ids, we calculate cosine similarity between the extracted sentences and the ground truth claims. We experiment with various similarity threshold to calculate TP, FP and FN, and set the final threshold to 0.95 to report best performing results. Finally, these metrics are used to report Precision@1, Recall@1 and F-1 scores.

5.3 Results

Table-3 shows the performance of baseline (CDC) (Levy et al., 2014) and three variants of MA-CIN models. We report two baselines - (1) spacy, and (2) Levy et al. (2014) using supervised learning method on CDC dataset which contains annotated claims. Since, Levy et al. (2014) do not report Recall and F1 scores, we have reported their Precision@1 score in this paper. We also train MA-CIN models on this dataset by removing all the annotations for self-supervised training. We observe that:

1. The combined variant of our self-supervised approach performs slightly better than the baseline on the CDC dataset. This shows that, MA-CIN models are able to learn similar properties as the baseline but without any sentence-level annotations. Thus, this eliminates the need to have an annotated dataset for claim identification.
2. MA-CIN models give comparable results on all three datasets. This shows the scalability of the models to identify “claim-worthy” sentences from any given article.
3. The combined variant of MA-CIN, which generates both the headline vector and the word in headline, performs better on all the datasets, except one: MA-CIN (OHVW) model trained on DNF-300 and evaluated on CDC_Eval performs slightly better than the combined model, however, the difference in the performance is very small.

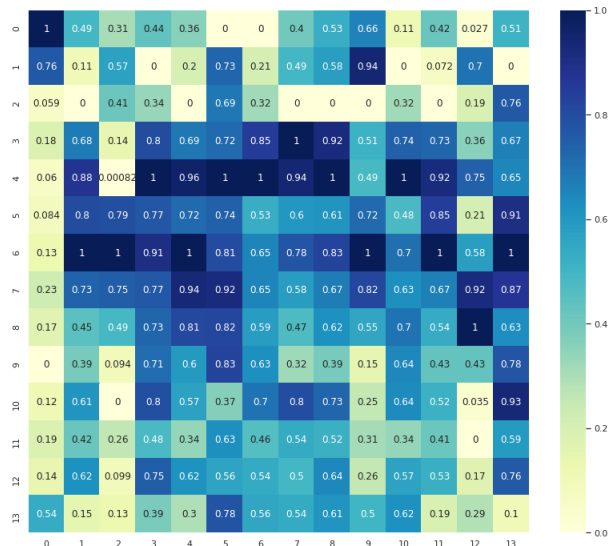


Figure 4: Interpretation of sentence-to-sentence relevance through attention weights.

6 Discussion

6.1 Analyzing Attention Weights

Attention weights help make the model interpretable to the end users by depicting relationship between all sentences as well as with the headline. From Figure-3, we can see that out of the top-5 predicted claims, 3 of them are present in the human evaluated test set. The last column, which contains attention coefficients between the headline and each sentence, depicts some interesting results -

(i) based on the human evaluation, the sentence having the least relevance with the headline is sentence 1. While this sentence contains words also present in the headline, the underlying meaning is not the same. This has been successfully captured by MA-CIN model by predicting sentence 1 as the least relevant claim;

(ii) further, highly ranked sentences 2, 12, and 13 have been correctly predicted as relevant claims by the model. This shows the model’s ability in learning the semantic relationship between the headline and the content of the article, and subsequently putting importance on sentences that are relevant to the headline’s underlying meaning. This property, which is also called “*aboutness*”, is efficiently exhibited by the model.

(iii) sentence 3, which is predicted by MA-CIN model as relevant with a score of 0.73, is not present in the ground truth. This indicates that the two annotators did not agree to have this sentence verified, even if it is relevant to the point of the

article. To analyze this further, we plan to conduct user studies as one of the future avenues.

(iv) sentence 4 is also predicted as a relevant claim but it's missing from the ground truth since the annotators did not agree to have this verified. The reason for this prediction could be because self-attention is able to identify the premise of highly relevant sentences. Hence, sentence 4, which is the continuation of highly relevant sentence 3, is also given importance by the headline. This relevance between sentences 3 and 4 is depicted in Figure-4, where the attention weight between these two is the highest.

From Figure-4, we also observe that:

(i) sentence 4 is highly relevant to sentences 3 to 8, which is intuitive, since the story of the intern forms the premise of the claims in the article;

(ii) sentences 2 and 4 have been shown to have the least relevance with each other which is also true as shown in Figure-3. The two sentences, if considered in isolation, make two different claims which are not related to each other;

(iii) the model has made sure that a sentence does not assign high relevance to itself as it would be counter-intuitive.

6.2 Limitations

Since, the evaluation methodologies for CDC dataset has not been explained clearly, in our paper, we have considered vector cosine similarity between the ground truth claim in the CDC_Eval and the extracted claim from the model which may leave a margin of error in the evaluation scores. Additionally, ground truth in DNF_Eval is manually generated and may contain subjective biases. Although these biases have been overcome by MACIN models, as explained in 6.1, but we also plan to enhance the ground truth judgement using crowdsourced annotation. We intend to use these annotations to fine-tune the models.

7 Conclusion and Future Work

In this work, we build a novel, self-supervised approach to identify “claim-worthy” sentences - an important task for automated fact checking. The focus of this work is on fake news articles where there is a deliberate intent to influence people or cause social unrest. We have introduced novel datasets of such articles with features essential for the downstream task of fake news verification. Using powerful attention models, we explore the notion of

aboutness of the headline and the content of the article to identify “claim-worthy” sentences. Experiments with three datasets show the strength of our model architecture in overcoming human-induced biases, which is quite common when using sentence-level claim-annotated datasets. Based on the comparison with the baseline, which was implemented using annotated dataset, we show that our models do not require annotated claims for training to identify claim-worthy sentences efficiently. We have also showed that our model is scalable to any dataset with topic and content.

Future work involves increasing the robustness of the models presented in this paper. We plan to use crowdsourced annotation on the dataset released with this paper to measure the *influence* of the article on general readers and then use these indicators to fine-tune our models. Experimentation with more robust sentence encoders is another avenue of future work. Additionally, going forward, we plan to identify a maximum of 3 claims per article which will be used for evidence-based fake news detection. We also plan to expand the dataset, presented in this work, to include fake news articles on topics other than Politics and Health.

References

- Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim. 2014. [A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics](#). In *Proceedings of the First Workshop on Argumentation Mining*, pages 64–68, Baltimore, Maryland. Association for Computational Linguistics.
- Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.
- Leif Azzopardi, Gabriella Kazai, Stephen Robertson, Stefan Rüger, Milad Shokouhi, Dawei Song, and Emine Yilmaz. 2009. *Advances in Information Retrieval Theory: Second International Conference on the Theory of Information Retrieval, ICTIR 2009 Cambridge, UK, September 10-12, 2009 Proceedings*, volume 5766. Springer.
- PD Bruza and Theo WC Huibers. 1996. A study of aboutness in information retrieval. *Artificial Intelligence Review*, 10(5-6):381–407.
- Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684. ACM.

- M de Cock Buning. 2018. *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation*. Publications Office of the European Union.
- Judith Eckle-Kohler, Roland Kluge, and Iryna Gurevych. 2015. [On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2236–2242, Lisbon, Portugal. Association for Computational Linguistics.
- Pepa Gencheva, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, and Ivan Koychev. 2017. [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*, pages 267–276, Varna, Bulgaria. INCOMA Ltd.
- D Graves. 2018. Understanding the promise and limits of automated fact-checking.
- Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. 2017. [Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster](#). In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1803–1812. ACM.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. [Long short-term memory](#). *Neural Computation*, 9(8):1735–1780.
- Mercedes Jaime Sis and MERCEDES. 2009. Titles or headlines? anticipating conclusions in biomedical research article titles as a persuasive journalistic strategy to attract busy readers. *Miscelnea: A Journal of English and American Studies*, 39:29–51.
- Israa Jaradat, Pepa Gencheva, Alberto Barrón-Cedeño, Lluís Màrquez, and Preslav Nakov. 2018. [Claim-Rank: Detecting check-worthy claims in Arabic and English](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 26–30, New Orleans, Louisiana. Association for Computational Linguistics.
- Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- Ron Kohavi. 1995. [A study of cross-validation and bootstrap for accuracy estimation and model selection](#). In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95*, pages 1137–1143, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Lev Konstantinovskiy, Oliver Price, Mevan Babakar, and Arkaitz Zubiaga. 2018. [Towards automated factchecking: Developing an annotation schema and benchmark for consistent automated claim detection](#). *arXiv preprint arXiv:1809.08193*.
- Jeffrey Kuiken, Anne Schuth, Martijn Spitters, and Maarten Marx. 2017. [Effective headlines of newspaper articles in a digital environment](#).
- Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. [Prominent features of rumor propagation in online social media](#). In *2013 IEEE 13th International Conference on Data Mining*, pages 1103–1108. IEEE.
- Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. [Gradient-based learning applied to document recognition](#). *Proceedings of the IEEE*, 86(11):2278–2323.
- Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni, and Noam Slonim. 2014. [Context Dependent Claim Detection](#). In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1489–1500, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.
- Ran Levy, Shai Gretz, Benjamin Sznajder, Shay Hummel, Ranit Aharonov, and Noam Slonim. 2017. [Un-supervised corpus-wide claim detection](#). In *ArgMining@EMNLP*.
- Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. [Detect rumors using time series of social context information on microblogging websites](#). In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15*, pages 1751–1754, New York, NY, USA. ACM.
- Jeppe Norregaard, Benjamin D. Horne, and Sibel Adali. 2019. [NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles](#). *CoRR*, abs/1904.01546.
- Archita Pathak and Rohini Srihari. 2019. [BREAK-ING! presenting fake news corpus for automated fact checking](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 357–362, Florence, Italy. Association for Computational Linguistics.
- Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. [Automatic detection of fake news](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3391–3401, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. [A stylistometric inquiry into hyperpartisan and fake news](#). *arXiv preprint arXiv:1702.05638*.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937.

Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. [Csi: A hybrid deep model for fake news detection](#). In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17*, pages 797–806, New York, NY, USA. ACM.

Sneha Singhania, Nigel Fernandez, and Shrisha Rao. 2017. 3han: A deep neural network for fake news detection. In *International Conference on Neural Information Processing*, pages 572–581. Springer.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention Is All You Need](#). *arXiv:1706.03762 [cs]*. ArXiv: 1706.03762.

Karin Wahl-Jorgensen and Thomas Hanitzsch. 2009. *The Handbook Of Journalism Studies*. Technical report.

William Yang Wang. 2017. ”liar, liar pants on fire”: A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*.

Liang Wu and Huan Liu. 2018. [Tracing fake-news footprints: Characterizing social media messages by how they propagate](#). In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18*, pages 637–645, New York, NY, USA. ACM.

Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. 2012. Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, page 13. ACM.

Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. 2018. [Self-attention generative adversarial networks](#).

A Appendices

A.1 Definitions

Fake News: Articles where there is a deliberate intent to influence readers through fabricated or manipulated claims in the headline and the content. Such articles have a compelling writing style similar to the mainstream media.

“Claim-worthy”: Statements which are important to the point of the article but one would require to have them verified.

Compelling Fake News Articles: Articles which make persuasive claims in headline and content,

that may influence readers to believe a fabricated/manipulated story.

Credible Sources: Mainstream media, established fact-checking websites and Government documents.

Questionable Sources: Non-mainstream media like infowars, naturalnews, Breitbart etc.

A.2 Experiment Architectures

A.2.1 Vector Generation

Architecture setting for generating Headline Vector (HV) displayed in Figure-5

input	Model	Output size	[Kernel size, Filters, Strides], Repeats	
sentence vectors		500x300		
	conv1d_1	500 x 256	[4, 256, 1] x 1	
	conv1d_2	500 x 512	[4, 256, 1] x 1	
	SelfAttention	500 x 512	[1, 64, 1] x 4	
	Concat	500 x 2048		
	conv1d_3	500 x 512	[4, 256, 1] x 1	
	headline vector		1 x 300	
		conv1d_4	1 x 512	[1, 512, 1] x 1
CrossAttention		500 x 512	[1, 64, 1] x 4	
Concat		500 x 2048		
conv1d_5		250 x 512	[4, 512, 2] x 1	
conv1d_6		125 x 512	[4, 512, 2] x 1	
conv1d_7		63 x 512	[4, 512, 2] x 1	
conv1d_8		32 x 512	[4, 512, 2] x 1	
Global Pooling	512			
FC_1	1024			
output_vector	300			

Figure 5: Architecture setting for generating Headline Vector(HV).

A.2.2 Word Generation

Architecture setting for generating Headline Vector Word Probabilities (OHVW) displayed in Figure-6

A.3 DNF-700 Dataset Details

Each article is identified by an *id*. The content of the article is stored in a separate text files having file name “article_*id*”, for example, article_122. A JSON file is also provided with the following fields:

id: Unique identifier of the article starting from 0.

authors: Authors of the article.

headline: Headline of the article.

type: “fake” (articles from Stanford and Buzzfeed datasets which are already proven fake); and “questionable” (articles from *Getting Real About Fake News* Kaggle dataset which require manual

input	Model	Output size	[Kernel size, Filters, Strides], Repeats
sentence vectors		500x300	
	conv1d_1	500 x 256	[4, 256, 1] x 1
	conv1d_2	500 x 512	[4, 256, 1] x 1
	SelfAttention	500 x 512	[1, 64, 1] x 4
	Concat	500 x 2048	
	conv1d_3	500 x 512	[4, 256, 1] x 1
headline vector		1 x 300	
	conv1d_4	1 x 512	[1, 512, 1] x 1
	CrossAttention	500 x 512	[1, 64, 1] x 4
	Concat	500 x 2048	
	conv1d_5	250 x 512	[4, 512, 2] x 1
	conv1d_6	125 x 512	[4, 512, 2] x 1
	conv1d_7	63 x 512	[4, 512, 2] x 1
	conv1d_8	32 x 512	[4, 512, 2] x 1
	Global Pooling	512	
	Repeat	50 x 512	
	Bi-LSTM	50 x 1024	
	TimeDistributed Dense, softmax	50 x 20000	

Figure 6: Architecture setting for generating Headline Vector Word Probabilities (OHVV).

verification of the degree of veracity)

urls: Source/domain URL of the article.

A.4 DNF-300 Dataset Details

DNF-300 is more sophisticated subset of DNF-700 with additional fields based on manual verification of the article. The JSON file of this dataset contains following fields:

id: Unique identifier of the article starting from 0.

authors: Authors of the article.

headline: Headline of the article.

type: {(0) False; (1) Partial Truth; (2) Opinions Stated As Fact; (3) True}

urls: Source/domain URL of the article.

evidence: URL of credible sources supporting or refuting the article. This field is empty when no evidence were found which talked about the claims made in this article. This means, the claims are innovated lies. In such cases, the *type* field is set as 0.

reason: Reason about the verdict. It can be one of the following:

1. Based on Snopes rating 'False' which means 'the primary elements of a claim are demonstrably false.'
2. Based on Snopes rating 'Unproven' which means 'insufficient evidence exists to establish the given claim as true, but the claim cannot be definitively proved false.'
3. Based on Snopes rating 'Mixture' which

means 'a claim has significant elements of both truth and falsity to it such that it could not fairly be described by any other rating.'

4. Based on Snopes rating 'Mostly False' which means 'the primary elements of a claim are demonstrably false, but some of the ancillary details surrounding the claim may be accurate.'
5. The key claim is false (based on Snopes rating), however, the article also contains opinions stated as fact.
6. Snopes mentions that a true story was manipulated to mislead people.
7. The key claims are true but exaggerated by adding personal opinions stated as fact.
8. No reports from trusted sources for the key claims.
9. True story manipulated to mislead readers by making unverifiable claims such as '*some_claim*'.
10. Article is fraught with opinions stated as fact about a true event.
11. Found evidence to refute key claims.
12. Article contains opinions stated as fact.
13. Evidence found to support key claims.

Topic:	The sale of violent video games to minors should be banned	
S1	Violent video games can increase children's aggression	V
S2	Video game addiction is excessive or compulsive use of computer and video games that interferes with daily life	X
S3	Many TV programmers argue that their shows just mirror the violence that goes on in the real world	X
S4	Violent video games should not be sold to children	X
S5	Video game publishers unethically train children in the use of weapons	V

Figure 7: : Example for CDCs and for statements that should not be considered as CDCs. The V and X indicate if the candidate is a CDC for the given Topic, or not, respectively.

A.5 Examples

We present examples for all 4 label types {*False*; *Partial Truth*; *Opinion stated as fact*; *True*} present in our dataset: DNF-300. Please refer Table-4,5,6,7. An annotated example from CDC dataset is displayed in Figure-7

Headline: Clinton Received Debate Questions Week Before Debate

The first presidential debate was held and Hillary Clinton was proclaimed the winner by the media. Indeed Clinton was able to turn in a strong debate performance, but did she do so fairly? Multiple reports and leaked information from inside the Clinton camp claim that the Clinton campaign was given the entire set of debate questions an entire week before the actual debate. Earlier last week an NBC intern was seen hand delivering a package to Clintons campaign headquarters, according to sources. The package was not given to secretarial staff, as would normally happen, but the intern was instead ushered into the personal office of Clinton campaign manager Robert Mook. Members of the Clinton press corps from several media organizations were in attendance at the time, and a reporter from Fox News recognized the intern, but said he was initially confused because the NBC intern was dressed like a Fed Ex employee. The reporter from Fox questioned campaign staff about the intern, but campaign staff at first claimed ignorance and then claimed that it was just a Fed Ex employee who had already left. No reporters present who had seen the intern dressed as a Fed Ex employee go into Mooks office saw him leave by the same front entrance. The Fox reporter who recognized the intern also immediately looked outside of the campaign headquarters and noted that there were no Fed Ex vehicles parked outside. Clinton seemed to have scripted responses ready for every question she was asked at the first debate. She had facts and numbers memorized for specific questions that it is very doubtful she would have had without being furnished the questions beforehand. The entire mainstream media has specifically been trying to portray Trump as a racist and a poor candidate. By furnishing Clinton with the debate questions NBC certainly hoped to make Clinton appear much more knowledgeable and competent than Trump. And though it is unlikely that anyone will be able to conclusively prove that Clinton was given the debate questions, it seems both logical and likely.

Type: 0 (*False*)**Authors:** Baltimore Gazette**URLs:** <http://www.freemarketcentral.com/index.php/post/2503/report-clinton-received-debate-questions-a-week-before-debate>

Evidence: [<https://www.snopes.com/fact-check/clinton-received-debate-questions-week-before-debate/>, <https://www.truthorfiction.com/hillary-clinton-received-debate-questions-advance/>]

Reason: Based on Snopes rating 'False' which means 'the primary elements of a claim are demonstrably false.'

Table 4: An example on *False* type from DNF-300 dataset.

Headline: Allergens in Vaccines Are Causing Life-Threatening Food Allergies

It would probably surprise few people to hear that food allergies are increasingly common in U.S. children and around the world . According to one public health website , food allergies in children aged 0-17 in the U.S. increased by 50% from 1997 to 2011. Although food allergies are now so widespread as to have become almost normalized, it is important to realize that millions of American children and adults suffer from severe rapid-onset allergic reactions that can be life-threatening. Foods represent the most common cause of anaphylaxis among children and adolescents. The United Kingdom has witnessed a 700% increase in hospital admissions for anaphylaxis and a 500% increase in admissions for food allergy since 1990. The question that few are asking is why life-threatening food allergies have become so alarmingly pervasive. A 2015 open access case report by Vinu Arumugham in the Journal of Developing Drugs , entitled “ Evidence that Food Proteins in Vaccines Cause the Development of Food Allergies and Its Implications for Vaccine Policy ,” persuasively argues that allergens in vaccines and specifically food proteins may be the elephant in the room. As Arumugham points out, scientists have known for over 100 years that injecting proteins into humans or animals causes immune system sensitization to those proteins. And, since the 1940s, researchers have confirmed that food proteins in vaccines can induce allergy in vaccine recipients. Arumugham is not the first to bring the vaccine-allergy link to the public's attention. Heather Fraser makes a powerful case for the role of vaccines in precipitating peanut allergies in her 2011 book, *The Peanut Allergy Epidemic: What's Causing It and How to Stop It*.

Type: 1 (*Partial Truth*)**Authors:** Admin - Orissa**URLs:** galacticconnection.com

Evidence: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3890451/>

Reason: The key claim is written in such a way so that it misleads people in thinking all the food related allergies in US are caused by vaccines. Found evidence which says these type of allergies are rare.

Table 5: An example on *Partial Truth* type from DNF-300 dataset.

Headline: George Soros: Trump Will Win Popular Vote by a Landslide but Clinton Victory a 'Done Deal'

In recent weeks, Democrats have attempted to paint Republican presidential nominee Donald J. Trump as a lunatic for claiming that the election is going to be rigged in favor of his Democratic rival, Hillary Clinton. Even Republican politicians and former politicians are telling Trump to knock off such talk. But, as usual, Trump's shrewdness and defiance of standard political decorum in which the "opposition" party merely rolls over and surrenders in the face of Democratic pressure is winning the day. None other than billionaire investor and longtime Democratic supporter George Soros has said that the fix is literally in for the election, in favor of Clinton no matter how much of the popular vote, and from which battleground states, Trump captures. As reported by Top Right News and other outlets, during a recent interview with Bloomberg News, Soros a Democrat mega-donor openly admitted that Trump will win the popular vote in a "landslide." However, he said that none of that would matter, because a President Hillary Clinton is already a "done deal." In the interview, which is now going viral, Soros says with certainty that Trump will take the popular vote, despite what the polls say now (which are completely rigged to oversample Democrats), but not the Electoral College, which will go to Clinton. When the reporter asks if that is already a "done deal" that Clinton will be our next president no matter what Soros says "yes," and nods his head. Is Soros just making a prediction out of overconfidence? Or does he truly know something most of us don't know?

Type: 2 (*Opinion Stated As Fact*)

Authors: J. D. Heyes

URLs: https://www.naturalnews.com/055789_George_Soros_Hillary_Clinton_electoral_college.html

Evidence: 1. <https://www.snopes.com/fact-check/george-soros-trump-will-win-popular-vote-by-a-landslide-but-clinton-victory-a-done-deal/>,
2. <https://www.bloomberg.com/news/videos/2016-01-22/soros-clinton-to-win-popular-vote-in-landslide>

Reason: The key claim is false (based on Snopes rating), however, the article also contains opinions stated as fact.

Table 6: An example on *Opinion stated as fact* type from DNF-300 dataset.

Headline: Donald Trump: Minnesota Has Suffered Enough Accepting Refugees

In a pitch to suspend the nations Syrian refugee program , Donald Trump said Minnesotans have "suffered enough" from accepting Somali immigrants into their state. "Here in Minnesota you have seen first hand the problems caused with faulty refugee vetting, with large numbers of Somali refugees coming into your state, without your knowledge, without your support or approval," Trump said at a Minneapolis rally Sunday afternoon. He said his administration would suspend the Syrian refugee program and not resettle refugees anywhere in the United States without support from the communities, while Hillary Clintons "plan will import generations of terrorism, extremism and radicalism into your schools and throughout your communities."

Type: 3 (*True*)

Authors: Henry Wolff

URLs: amren.com

Evidence: 1. <https://time.com/4560078/donald-trump-minnesota-somali-refugees/>,
2. <https://www.buzzfeednews.com/article/claudiakoerner/trump-vs-somali-refugees>

Reason: Evidence found to support key claims.

Table 7: An example on *True* type from DNF-300 dataset.