Exploring Summarization to Enhance Headline Stance Detection

Robiert Sepúlveda-Torres(✉), Marta Vicente, Estela Saquete, Elena Lloret, and Manuel Palomar

Department of Software and Computing Systems, University of Alicante, Carretera de San Vicente s/n 03690, Alicante, Spain
{rsepulveda,mvicente,stela,elloret,mpalomar}@dlsi.ua.es

Abstract. The spread of fake news and misinformation is causing serious problems to society, partly due to the fact that more and more people only read headlines or highlights of news assuming that everything is reliable, instead of carefully analysing whether it can contain distorted or false information. Specifically, the headline of a correctly designed news item must correspond to a summary of the main information of that news item. Unfortunately, this is not always happening, since various interests, such as increasing the number of clicks as well as political interests can be behind of the generation of a headlines that does not meet its intended original purpose. This paper analyses the use of automatic news summaries to determine the stance (i.e., position) of a headline with respect to the body of text associated with it. To this end, we propose a two-stage approach that uses summary techniques as input for both classifiers instead of the full text of the news body, thus reducing the amount of information that must be processed while maintaining the important information. The experimentation has been carried out using the Fake News Challenge FNC-1 dataset, leading to a 94.13% accuracy, surpassing the state of the art. It is especially remarkable that the proposed approach, which uses only the relevant information provided by the automatic summaries instead of the full text, is able to classify the different stance categories with very competitive results, so it can be concluded that the use of the automatic extractive summaries has a positive impact for determining the stance of very short information (i.e., headline, sentence) with respect to its whole content.

Keywords: Natural language processing · Fake news · Misleading headlines · Stance detection

1 Introduction

Headlines are fundamental parts of news stories, summarizing the content and giving the reader a clear understanding of the article's content [9]. However, nowadays, the speed at which information spreads and the degree of information overload are considered by many to be reaching an unmanageable state [34].

© Springer Nature Switzerland AG 2021
E. Métais et al. (Eds.): NLDB 2021, LNCS 12801, pp. 243–254, 2021.
https://doi.org/10.1007/978-3-030-80599-9_22

Therefore, it is tempting to read only the headlines of news and share it without having read the entire story [15]. In this sense, a headline should be as effective as possible, without losing accuracy or being misleading [19], in order to maintain accuracy and veracity of the entire article.

Unfortunately, in practice, headlines tend to be more focused on attracting the reader's attention and going viral because of this, despite the lack of veracity within the information in the body text, thus leading to mis- or disinformation through erroneous/false facts or headline/body dissonance [6]. Headlines are considered misleading or incongruent when they significantly misrepresent the findings reported in the news article [7], by exaggerating or distorting the facts described in the news article. Some important nuances that are part of the news body text are missing in the headline, causing the reader to come to the wrong conclusion. Therefore, the reader cannot discover these inconsistencies if the news body text is not read [38].

In the research community, the task of automatically detecting misleading/incongruent headlines is addressed as a stance detection problem, which implies estimating the relative perspective, i.e., the stance of two pieces of text relative to a topic, claim or issue [14]. This is done through news body text analysis, determining the evidences from which the headline has been derived.

In this context, the main objective of our research is to propose a novel approach that automatically determines the stance of the headline with respect to its body text integrating summarization techniques in a two-stage classification problem, where both the news headline and its corresponding body text are given as input.

2 Related Work

Triggered by a greater demand for new technologies together with an increase in the availability of annotated corpora, headline stance detection task quickly emerged in the context of fake news analysis. In this context, research challenges and competitions, such The *Fake News Challenge*[1] (FNC-1) [2] were proposed.

FNC-1 was created using Emergent dataset [14] as a starting point [31] and it aimed to compile a gold standard to explore Artificial Intelligence technologies, especially ML and Natural Language Processing (NLP), applied to detection of fake news. The three best systems in this competition were Talos [3], Athene system [1] and UCLMR [30] in this order. Talos [3] applied a one-dimensional convolution neural networks (CNN) on the headline and body text, represented at the word level using Google News pretrained vectors. The output of this CNN is then sent to a multilayer perceptron (MLP) with 4-class output: *agree*, *disagree*, *discuss*, and *unrelated*, and trained end-to-end. Using this combination CNN-MLP, the system outperformed all the submissions and achieved the first position in the FNC-1 challenge. Outside the FNC-1 competition but using its dataset other work and experiments have been carried out. [40] addressed the

[1] http://www.fakenewschallenge.org/ (accessed online 18 March, 2021).

problem proposing a hierarchical representation of the classes, which combines agree, disagree and discuss in a new related class. A two-layer neural network is learning from this hierarchical representation of classes and a weighted accuracy of 88.15% is obtained with their proposal. Furthermore, [12] constructed a stance detection model by performing transfer learning on a RoBERTa deep bidirectional transformer language model by taking advantage of bidirectional cross-attention between claim-article pairs via pair encoding with self-attention. They reported a weighted accuracy of 90.01%. Outside the FNC-1 Challenge and dataset, there is other research that also addresses the stance detection tasks, determining the relation of a news headline with its body text by extracting key quotes [28] or claims [36].

Turning now into text summarization, its main potential is its ability to extract the most relevant information from a document, and synthesize its essential content. In this respect, one of the most outstanding areas in using summarization techniques is that of news, partly thanks to the development of appropriate corpora (e.g. DUC, Gigaword, CNN/DailyMail)[8], and the wide range of techniques and approaches to help digest this type of information [11,22,26,41]. Moreover, there is a significant amount of research on the task of headline generation using summarization techniques [4,10,39], and more recently using Deep Learning [16,18,33].

However, to the best of our knowledge, regarding disinformation, summarization for detecting fake news has only been proposed in [13], where an abstractive summarization model is applied. In this manner, the news article is first summarized, and the generated summary is used by the classification algorithm instead of the whole body text, which may be too long, or just the headline, which may be too short. Considering this aforementioned research results in which the accuracy is higher when using the summary compared to the full body text, our approach adopts this similar idea where the news article is reduced to its essential information, and exploits it further within a two-stage classifier to detect incongruities between headline and the body text of a news article.

3 Approach Architecture

Following the FNC-1 guidelines, the task of detecting misleading headlines tackled as a headline stance detection task involves classifying the stance of the body text with respect to the headline into one of the following four classes: a) *agrees*—agreement between body text and headline; b) *disagrees*—disagreement between body text and headline; c) *discusses*—same topic discussed in body text and headline, but no position taken; and, d) *unrelated*—different topic discussed in body text and headline.

To address this task, we propose an approach[2] that involves two-stages, thus addressing the task as a two-level classification problem: *Relatedness Stage*, and *Stance Stage*. Figure 1 illustrates the complete architecture. Next, a more

[2] Implementation available at https://github.com/rsepulveda911112/Headline-Stance-Detection.

detailed description of both stages and the different modules involved in performing the stance classification is provided.

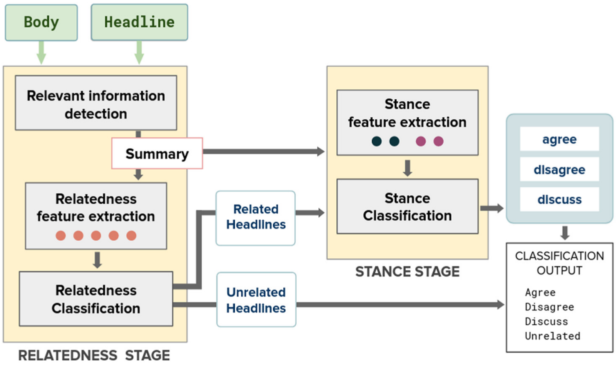

Fig. 1. Two-staged architecture devised to tackle the headline stance detection task

3.1 Relatedness Stage

The *Relatedness Stage* is in charge of determining whether or not the headline and the body of the news are related. The inputs of this stage are both the text body and the headline, resulting in a binary classification. The outputs of this stage are:

– The *headlines* classified as *related* or *unrelated*.
– The *summary* of the news content, obtained in a relevant information detection module.

For this, three modules are proposed: i) relevant information detection; ii) relatedness feature extraction; and, iii) relatedness classification.

Relevant Information Detection Module. This module aims to create a summary revealing the important information of the input news article in relation to its headline. Although different summarization approaches could be used for this purpose, we opt for the popular and effective TextRank extractive summarization algorithm [24], due to its good performance, execution time and implementation availability.[3] This algorithm represents the input text as a graph, where the vertices represent the sentences to be ranked, and the edges are the connections between them. Such connections are determined by the similarity between the text sentences measured with respect to their overlapping content. Then, a weight is computed for each of the graph edges indicating the strength of the connection between the sentences pairs/vertices linked by them. Once the

[3] https://pypi.org/project/sumy/.

graph is built, a weighted graph-based ranking is performed in order to score each of the text sentences. The sentences are then sorted in reversed order of their score. Finally, the top ranked sentences, in our case five, are selected to be included in the final summary.

Relatedness Feature Extraction. This module is focused on computing similarity metrics between the generated summary and the given headline. The computed features, which will be used in subsequent module, are:

– Cosine similarity between headline and summary TF-IDF vector without stop word [27].
– Overlap coefficient between headline and summary without stop words [23].
– BERT cosine similarity between headline and summary. We use sentences transformer [29].
– Positional Language Model (PLM) salience score between headline and summary, which has been shown to be effective for relevant content selection [35]
– Soft cosine similarity between headline and summary without stop words. We use word2vec vector [25].

Relatedness Classification. This module exploits the relatedness features previously computed, as well as the automatic summary to finally classify the headlines as *related* or *unrelated*. The proposed architecture is flexible to choose any model that allows classifiers to be improved.

In this case, the design of the relatedness classification module is based on fine-tuning the RoBERTa (Robustly optimized BERT approach) pre-trained model [21], applying a classifier to its output afterwards.

First, the headline and the summary are concatenated and processed with the RoBERTa model. The resulting vector is consecutively multiplied by the three features (Cosine similarity, Overlap coefficient, BERT cosine similarity, PLM salience score and Soft cosine similarity) to finally carry out the classification using a Softmax activation function in the output layer.

Specifically, we have chosen RoBERTa Large model (24 layer and 1024 hidden units) since it achieves state-of-the-art results in General Language Understanding Evaluation (GLUE) [37], Reading Comprehension Dataset From Examinations (RACE) [20] and Stanford Question Answering Dataset (SQuAD) benchmark. Similar to [12,21,32], in this work we fine-tune RoBERTa to efficiently address a task that involves comparing sentences.

In our model, the hyperparameter values are: maximum sequence length of 512; batch size of 4; training rate of 1e-5; and, training performed for 3 epochs. These values were established after successive evaluations, following previous experiments on this model [12,21,32].

3.2 Stance Stage

Once our approach has been able to identify the headlines that are related to their source text, the main goal of this stage is to determine their type considering the remaining stances: *agree*, *disagree* or *discuss*. Therefore, the claim made in the headline can be finally classified into one of three classes left.

The inputs of this stage are:

- The *headlines* classified as *related*.
- The *summary* of the news content.

These classified headlines together with the *unrelated* headlines determined before, will comprise the final output for the whole approach. To achieve this, this stage comprises the following modules:

Stance Feature Extraction. In this module, polarity features of the headline and the summary are computed using NLTK tool [5].

- Polarity positive and negative of the headline (Pol_head_pos, Pol_head_neg).
- Polarity positive and negative of the summary (Pol_sum_pos, Pol_sum_neg).

Stance Classification. Similar to the Relatedness classification module, this stage has been build using RoBERTa as foundation, selected as the model able to improve the classification. In this case, the four features of the stance feature extraction module are added, two dense layers are included to reduce dimensions and, finally, the Softmax classification layer. The hyperparameters of the model used in this classifier are the same as those of the Relatedness classification, except for the classification output which in this case is of three classes: *agree*, *disagree*, *discuss*. In all this classification process, the automatic summaries previously generated with TextRank are used.

4 Evaluation and Discussion

The evaluation of our proposed approach is conducted over the Fake News Challenge dataset (FNC-1) whose instances are labeled as *agree*, *disagree*, *discuss* and *unrelated*. The dataset contains 1,683 news with their headlines and was split into a training set (66.3%) and a testing set (33.7%), where neither the headlines nor the body text overlapped.

To measure our approach's performance, a set of incremental experiments were conducted, where each of the two stages of the proposed architecture were first evaluated independently, and then, the whole approach was validated. By this means, we can first measure the effectiveness of this stage in isolation, also conducting an ablation study to verify whether or not the features used in each of the stages of the classifier make a positive contribution.

In addition to the average accuracy and *Relative Score* metric originally proposed in the FNC-1 challenge,[4] we also take into account the F_1 class-wise,

[4] This metric assigns higher weight to examples correctly classified, as long as they belonged to a different class from the *unrelated* one.

and a macro-averaged F_1[5] (F_1m) metrics [17]. The advantage of these additional metrics is that it is not affected by the size of the majority class.

Table 1 shows the performance obtained in *Relatedness Stage* (first classifier).

Table 1. Relatedness classification results using automatic summaries

System	F_1 Score		F_1m
	Related	Unrelated	
Relatedness Stage FNC-1-Summary	98.22	99.31	98.77

The ablation study for this stage consisted on performing five different experiments removing each time one specific feature with the aim of gain better insights on how each of these features contribute to the proposal. Results are shown in Table 2 and indicate that the most influential feature for the classification is the Cosine similarity since the experiment that does not use this feature obtains the worst results, although the classification results are still very high.

Table 2. Ablation study results for the features used in the *Relatedness Stage*

Removed feature	F_1 Score		F_1m
	Related	Unrelated	
Cosine similarity	97.52	99.04	98.28
BERT cosine similarity	97.66	99.11	98.38
PLM salience score	97.91	99.19	98.55
Overlap coefficient	98.04	99.24	98.64
Soft cosine similarity	98.05	99.26	98.66

Concerning the validation of the *Stance Stage* in isolation, only the examples tagged as *related* from the FNC-1 Gold-Standard are used. Table 3 shows the performance results obtained in the *Stance Stage* (second classifier).

Table 3. *Stance Stage* results

Removed feature	F_1 Score			F_1m
	Agree	Disagree	Discuss	
Stance Stage FNC-1	74.54	64.54	87.69	75.59

As we did with the *Relatedness Stage*, an ablation study (Table 4) was carried out, where the *Stance Stage* classifier was tested removing each of the proposed

[5] This is computed as the mean of those per-class F scores.

features (Pol_head_pos, Pol_head_neg, Pol_sum_pos, Pol_sum_neg). The included features clearly show their positive influence in the performance of the classifier. In this case the most influential feature for the classification is the *Pol_head_pos*.

Table 4. Ablation study results for the features used in the *Stance Stage*

Removed feature	F_1 Score			F_1m
	Agree	Disagree	Discuss	
Pol_head_pos	71.64	56.99	87.10	71.91
Pol_head_neg	72.19	58.84	88.12	73.05
Pol_sum_neg	71.68	61.31	88.11	73.70
Pol_sum_pos	73.08	59.94	88.26	73.76

Finally, the results of the whole approach, which integrates the Relatedness and Stance classifiers together with the sole use of automatic summaries for these two classifiers are shown in Table 5. This table contains the performance for the class-wise F_1, macro-average F_1m, accuracy (Acc.) and the relative score (Rel. Score). Moreover, it also provides the results obtained by competitive state-of-the-art systems together with additional configurations that were also tested.

Table 5. Complete approach performance and comparison with state-of-the-art systems

System	F_1 Score				F_1m	Acc.	Rel. Score
	Agree	Disagree	Discuss	Unrelated			
Talos [3]	53.90	3.54	76.00	99.40	58.21	89.08	82.02
Athene [1]	48.70	15.12	78.00	99.60	60.40	89.48	82.00
UCLMR [30]	47.94	11.44	74.70	98.90	58.30	88.46	81.72
Human Upper Bound [1]	58.80	66.70	76.50	99.70	75.40	–	85.90
Dulhanty et al. [12]	73.76	55.26	85.53	99.12	78.42	93.71	90.00
Zhang et al. [40]	67.47	**81.30**	83.90	**99.73**	**83.10**	93.77	89.30
OurApproach-1stage	71.64	53.31	85.25	99.29	77.37	93.58	89.92
OurApproach-2stages	**74.22**	64.29	**86.00**	99.31	80.95	**94.13**	**90.73**

The 3 first rows are the top-3 best systems that participated in the FNC-1 challenge, calculated using the confusion matrices and results published [30] or made available by the authors.[6],[7]

The fourth row corresponds to the *Human Upper Bound* [1], and is the result of conducting the FNC-1 stance detection task manually.

[6] https://github.com/hanselowski/athene_system/ (accessed online 15 March, 2021).
[7] https://github.com/Cisco-Talos/fnc-1 (accessed online 15 March, 2021).

Next, the fifth and sixth rows include the results of recent approaches [12,40] that also addressed the headline stance detection task using the FNC-1 dataset, but did not participate in this challenge. Since there was no public code available, these results were also calculated from the confusion matrices provided in their respective papers.

The seventh row indicates the results for our approach but configured only as a single classifier (*OurApproach-1stage*). Finally, the last row belongs to our approach, using our proposed two-stage classification (*OurApproach-2stages*). Regardless whether the classification is conducted in 1 or 2 stages, both approaches use for the whole process the features extracted and the summaries created from the full body text.

As can be seen in Table 5, *OurApproach-2stages* is competitive enough with respect to the other systems, given that it only uses short summaries for the classification process, and not the full body text as the other systems use, so the information reduction does not imply a high loss in the results obtained, being better than the FNC-1 participants, and the human upper bound. Furthermore, the results also validate the fact that dividing the classification into two stages is beneficial and yields better performance with respect to using our proposed model with a single classifier (rows 7th and 8th), especially for detecting disagreement between the headline and the news article. At this point, it is worth noting that the results previously obtained with the independent evaluation of the *Stance Stage* are slightly better the ones of whole approach (see Table 3). This was already expected since errors derived from the *Relatedness Stage* were avoided in the former, simulating an ideal environment.

Whereas our approach outperforms the other automatic systems in terms of *agree* and *discuss* classes, accuracy, and relative score, it was outperformed in the *disagree* class by [40] and in the *unrelated* class by top-3 best systems that participated in the FNC-1 challenge and [40]. When the results obtained by the participants in the FNC-1 competition are analyzed independently for each of the classes, it can be seen that except for the classification of *unrelated* headlines —whose results are close to 100% in F1 measure, and this happens also for the remaining approaches as well— for the remaining classes, the results are very limited. The systems that participated in the FNC-1 competition have a very reduced performance especially in detecting the *disagree* stance, whereas the detection of *agree* is around 50% in F1 measure and for *discuss* around 75% for the best approach. Outside the FNC-1 competition, the performance increases in all categories, being the *disagree* category one of the most challenging to classify, in which only the approach proposed in [40] obtains surprisingly high results for this category compared to the remaining methods.

5 Conclusions and Future Work

This paper presented an approach for stance detection, i.t., for automatically determining the relation between a news headline and its body. Its novelty relies on two key premises: i) the definition of a two-stages architecture to tackle the stance classification problem; and ii) the use of summarization instead of the whole news body. To show the appropriateness of the approach, different experiments were carried out in the context of an existing task —Fake News Challenge FNC-1—, where the stance of a headline had to be classified into one of the following classes: *unrelated*, *agree*, *disagree*, and *discuss*. The experiments involved validating each of the proposed classification stages in isolation together with the whole approach, as well as a comparison with respect the state of the art in this task.

The results obtained by our system were very competitive compared to other systems obtaining 94.13% accuracy, as well as the highest result in FNC-1 relative score compared with the state of the art (90.73%). Given that the use of summaries provided good results in this preliminary research, as a future goal, we would like to study more in-depth the impact of the summarization techniques in the stance detection process, by using other summarization approaches, or analysing how the length of the summaries affect the performance of the approach, among other issues to be researched. Moreover, we also plan to include in our stance detection approach, new learning strategies and discourse aware techniques, with the final aim to help to combat online fake news, a societal problem that requires concerted action.

Acknowledgements. This research work has been partially funded by Generalitat Valenciana through project "SIIA: Tecnologias del lenguaje humano para una sociedad inclusiva, igualitaria, y accesible" (PROMETEU/2018/089), by the Spanish Government through project "Modelang: Modeling the behavior of digital entities by Human Language Technologies" (RTI2018-094653-B-C22), and project "INTEGER - Intelligent Text Generation" (RTI2018-094649-B-I00). Also, this paper is also based upon work from COST Action CA18231 "Multi3Generation: Multi-task, Multilingual, Multimodal Language Generation".

References

1. Hanselowski, A., Avinesh, P.V.S., Schiller, B., Caspelherr, F.: Description of the system developed by team athene in the FNC-1 (2017). https://github.com/hanselowski/athene_system/blob/master/system_description_athene.pdf
2. Babakar, M., et al.: Fake News Challenge - I (2016). http://www.fakenewschallenge.org/. Accessed 29 May 2020
3. Baird, S., Sibley, D., Pan, Y.: Talos targets disinformation with fake news challenge victory (2017). blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html. Accessed 29 May 2020
4. Banko, M., Mittal, V.O., Witbrock, M.J.: Headline generation based on statistical translation. In: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, pp. 318–325. Association for Computational Linguistics (2000)
5. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc. (2009)
6. Chen, Y., Conroy, N.J., Rubin, V.L.: News in an online world: the need for an "automatic crap detector". In: Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community. American Society for Information Science (2015)

7. Chesney, S., Liakata, M., Poesio, M., Purver, M.: Incongruent headlines: yet another way to mislead your readers. Proc. Nat. Lang. Process. Meets J. **2017**, 56–61 (2017)
8. Dernoncourt, F., Ghassemi, M., Chang, W.: A repository of corpora for summarization. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation. European Language Resources Association (2018)
9. van Dijk, T.A.: News as Discourse. L. Erlbaum Associates, Communication Series (1988)
10. Dorr, B., Zajic, D., Schwartz, R.: Hedge trimmer: a parse-and-trim approach to headline generation. In: Proceedings of the North American of the Association for Computational Linguistics, Text Summarization Workshop, pp. 1–8 (2003)
11. Duan, Y., Jatowt, A.: Across-time comparative summarization of news articles. In: Proceedings of the 12th ACM International Conference on Web Search and Data Mining, pp. 735–743. Association for Computing Machinery, New York (2019)
12. Dulhanty, C., Deglint, J.L., Daya, I.B., Wong, A.: Taking a stance on fake news: Towards automatic disinformation assessment via deep bidirectional transformer language models for stance detection. arXiv preprint arXiv:1911.11951 (2019)
13. Esmaeilzadeh, S., Peh, G.X., Xu, A.: Neural abstractive text summarization and fake news detection. Computing Research Repository abs/1904.00788 (2019)
14. Ferreira, W., Vlachos, A.: Emergent: a novel data-set for stance classification. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, pp. 1163–1168. Association for Computational Linguistics (2016)
15. Gabielkov, M., Ramachandran, A., Chaintreau, A., Legout, A.: Social clicks: what and who gets read on Twitter? ACM SIGMETRICS Performance Eval. Rev. **44**, 179–192 (2016)
16. Gavrilov, D., Kalaidin, P., Malykh, V.: Self-attentive model for headline generation. In: Azzopardi, L., Stein, B., Fuhr, N., Mayr, P., Hauff, C., Hiemstra, D. (eds.) ECIR 2019. LNCS, vol. 11438, pp. 87–93. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15719-7_11
17. Hanselowski, A., et al.: A retrospective analysis of the fake news challenge stance-detection task. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 1859–1874. Association for Computational Linguistics (2018)
18. Iwama, K., Kano, Y.: Multiple news headlines generation using page metadata. In: Proceedings of the 12th International Conference on Natural Language Generation, pp. 101–105. Association for Computational Linguistics (2019)
19. Kuiken, J., Schuth, A., Spitters, M., Marx, M.: Effective headlines of newspaper articles in a digital environment. Digit. J. **5**(10), 1300–1314 (2017)
20. Lai, G., Xie, Q., Liu, H., Yang, Y., Hovy, E.: RACE: large-scale reading comprehension dataset from examinations. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 785–794. Association for Computational Linguistics (2017)
21. Liu, Y., et al.: RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
22. Mackie, S., McCreadie, R., Macdonald, C., Ounis, I.: Experiments in newswire summarisation. In: Ferro, N., et al. (eds.) ECIR 2016. LNCS, vol. 9626, pp. 421–435. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30671-1_31
23. Metcalf, L., Casey, W.: Metrics, similarity, and sets. In: Cybersecurity and Applied Mathematics, pp. 3–22. Elsevier (2016)

24. Mihalcea, R., Tarau, P.: TextRank: bringing order into text. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp. 404–411. Association for Computational Linguistics (2004)
25. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
26. Nenkova, A.: Automatic text summarization of newswire: lessons learned from the document understanding conference. In: Proceedings of the 20th National Conference on Artificial Intelligence, vol. 3, pp. 1436–1441. AAAI Press (2005)
27. Passalis, N., Tefas, A.: Learning bag-of-embedded-words representations for textual information retrieval. Pattern Recogn. **81**, 254–267 (2018)
28. Pouliquen, B., Steinberger, R., Best, C.: Automatic detection of quotations in multilingual news. Proc. Recent Adv. Nat. Lang. Process. **2007**, 487–492 (2007)
29. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
30. Riedel, B., Augenstein, I., Spithourakis, G.P., Riedel, S.: A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. Computing Research Repository, CoRR abs/1707.03264 (2017)
31. Silverman, C.: Lies, damn lies and viral content (2019). http://towcenter.org/research/lies-damn-lies-and-viral-content/. Accessed 29 May 2020
32. Slovikovskaya, V.: Transfer learning from transformers to fake news challenge stance detection (FNC-1) task. arXiv preprint arXiv:1910.14353 (2019)
33. Tan, J., Wan, X., Xiao, J.: From neural sentence summarization to headline generation: a coarse-to-fine approach. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 4109–4115. AAAI Press (2017)
34. Tsipursky, G., Votta, F., Roose, K.M.: Fighting fake news and post-truth politics with behavioral science: the pro-truth pledge. Behav.Soc. Issues **27**(1), 47–70 (2018). https://doi.org/10.5210/bsi.v27i0.9127
35. Vicente, M.E., Pastor, E.L.: Relevant content selection through positional language models: an exploratory analysis. Proces. del Leng. Nat. **65**, 75–82 (2020)
36. Vlachos, A., Riedel, S.: Identification and verification of simple claims about statistical properties. Proc. Conf. Empirical Methods Nat. Lang. Process. **2015**, 2596–2601 (2015)
37. Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., Bowman, S.: GLUE: a multi-task benchmark and analysis platform for natural language understanding. In: Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, pp. 353–355. Association for Computational Linguistics (2018)
38. Wei, W., Wan, X.: Learning to identify ambiguous and misleading news headlines. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 4172–4178. AAAI Press (2017)
39. Zajic, D., Dorr, B., Schwartz, R.: Automatic headline generation for newspaper stories. In: Proceedings of the Workshop on Automatic Summarization 2002, pp. 78–85 (2002)
40. Zhang, Q., Liang, S., Lipani, A., Ren, Z., Yilmaz, E.: From stances' imbalance to their hierarchical representation and detection. In: The World Wide Web Conference, pp. 2323–2332. ACM (2019)
41. Zhu, C., Yang, Z., Gmyr, R., Zeng, M., Huang, X.: Make lead bias in your favor: A simple and effective method for news summarization. arXiv preprint arXiv:1912.11602 (2019)