Benchmarking Multi-Modal Entailment for Fact Verification

Parth Patwa[1], Shreyash Mishra[2], S Suryavardan[2], Amrit Bhaskar[3], Parul Chopra[4], Aishwarya Reganti[5], Amitava Das[6,7], Tanmoy Chakraborty[8], Amit Sheth[7], Asif Ekbal[9] and Chaitanya Ahuja[3]

[1]*University of California Los Angeles, USA*
[2]*IIIT Sri City, India*
[3]*Arizona State University, USA*
[4]*Carnegie Mellon University, USA*
[5]*Amazon, USA*
[6]*Wipro AI labs, India*
[7]*AI Institute, University of South Carolina, USA*
[8]*IIIT Delhi, India*
[9]*IIT Patna, India*

Abstract

Fake news can spread quickly on social media and it is important to detect it before it creates lot of damage. Automatic fact/claim verification has recently become a topic of interest among diverse research communities. We present the findings of the Factify shared task, which aims undertake multi-modal fact verification, organized as a part of the De-Factify workshop at AAAI'22. The task is modeled as a multi-modal entailment task, where each input needs to be classified into one of 5 classes based on entailment and modality. A total of 64 teams participated in the Factify shared task, and of them, 9 teams submitted their predictions on test set. The most successful models were BigBird or other variations of BERT. The highest F1 score averaged across all the classes was 76.82%.

Keywords

Fake News, Fact Verification, Multimodality, Dataset, Machine Learning, Entailment

1. Introduction

Checking facts is a time-consuming process. Due to the speed with which both knowledge and misinformation spreads in today's media ecosystem, fact-checking has become increasingly important. A wide range of stakeholders (represented by journalists, scholars, and independent fact checkers) have worked together to defend communities against incorrect information. A typical fact-checking process involves establishing the contested claims, seek expert opinions, collect relevant information, verify sources, checking missing information, debate, and then reach a conclusion. Therefore, manual fact-checking, while very accurate, is a time-consuming

De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, co-located with AAAI 2022. 2022 Vancouver, Canada

parthpatwa@g.ucla.edu (P. Patwa); shreyash.m19@iiits.in (S. Mishra); suryavardan.s19@iiits.in (S. Suryavardan); abhask10@asu.edu (A. Bhaskar); parulcho@andrew.cmu.edu (P. Chopra); amitava.das2@wipro.com (A. Das)

and tedious process. With the scale at which content is generated in the present day, it is practically impossible for human annotators to manually verify facts. As a result, many researchers have been looking into how fact-checking may be automated, employing techniques such as natural language processing, machine learning, knowledge representation, and databases to predict the veracity of claims automatically. The topic of automatic fact/claim checking has recently piqued the interest of many research communities. FEVER [1] and FNC [2] are two forums dedicated to discussing automatic text fact-checking.

Although there are research initiatives and datasets on textual fact verification [3, 4, 5, 6], there is less focus on multi-modal or cross-modal fact verification. Both image and text offer a wealth of information, but they do it in different ways as seen in these multimodal datasets [7, 8, 9]. When comparing representation learning within the same modality, the created model for cross-modal architecture must learn not only image and text features to convey their respective content, but also a measure for inter-modal relationships. In this new task, we look at multimodal entailment. The task is to detect multimodal fake news, where each data sample contains true information source and another source whose correctness is to be verified. The goal is to identify semantic and structural differences between real news pieces and fake ones.

This paper describes the details of shared task on multimodal fact verification, Factify [10], which was organized as part of the De-Factify workshop at AAAI 2022. Our work attempts to bring to light the importance of images in fact verification, and generate research interest for multimodal fact verification.

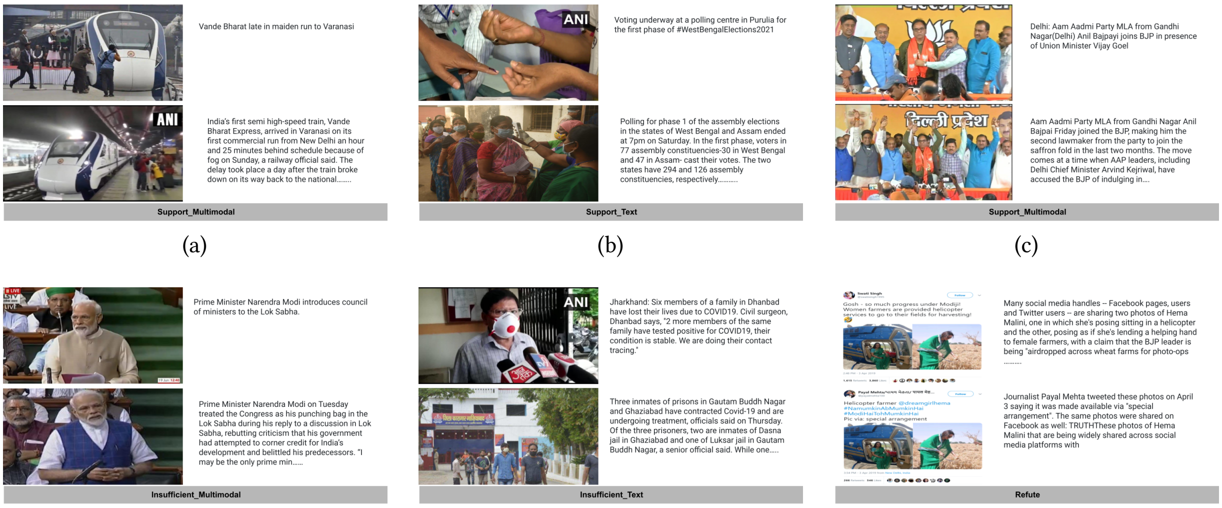
The paper is organized as follows, In the next section, we illustrate related work in the domain of automatic fact checking in regards to both datasets and approaches. In Section 3, the task details are provided including dataset statistics and the baseline models developed by us. In section 4 and **??**, we enlist the participating teams' approaches and results obtained respectively. In the last section we conclude with directions for future work.

2. Related Work

Misleading or false information passed as real is called fake news. Regardless of whether it is deliberate or not, repeated misinformation makes it difficult for us to discern and distinguish what's real [11]. Fake news can be in the form of conspiracies, satire, misinformation or pseudoscience. The presence of such unreliable information can affect mental health, cause panic and change people's beliefs, the COVID-19 pandemic being a fitting example [12] [13]. This vulnerability can be exploited to push political agendas, marketing etc. Moreover, social media escalates this issue by providing the world an open and ungoverned platform. Without mitigation, user's emotions, personal beliefs and lack of data can escalate the spread of fake news [14].

Despite people being aware of such misinformation, the spread of fake news can not be mitigated due to lack of information, time or religiosity [15]. The effort from fact checking and verification organisations is a great initiative but it is strenuous to manually verify the endless stream of social media data. As a result, the requirement for automated fact verification has been recognised by several researchers. Workshops and shared tasks like FEVER [1], Fakeddit [7], Constraint2021 [16], pan2020 [17], DeepFake challenge [18] etc. have drawn attention to this task. Along with FEVER, other datasets such as LIAR [3], CREDBANK [4], Constraint2021 [19] etc. have focused on fact verification of textual modality. On the other hand Fakeddit and the Whatsapp fact-checking dataset [9] provide multi-modal data. Some researchers have also worked towards presenting models and algorithms for this task using CNN, RNN, VAE, adversarial modeling, entailment etc. [20, 21, 22, 23, 24, 25, 26, 27, 28].

3. Task Details



Vande Bharat late in maiden run to Varanasi

India's first semi high-speed train, Vande Bharat Express, arrived in Varanasi on its first commercial run from New Delhi an hour and 25 minutes behind schedule because of fog on Sunday, a railway official said. The delay took place a day after the train broke down on its way back to the national........

Support_Multimodal

(a)

Voting underway at a polling centre in Purulia for the first phase of #WestBengalElections2021

Polling for phase 1 of the assembly elections in the states of West Bengal and Assam ended at 7pm on Saturday. In the first phase, voters in 77 assembly constituencies-30 in West Bengal and 47 in Assam- cast their votes. The two states have 294 and 126 assembly constituencies, respectively...........

Support_Text

(b)

Delhi: Aam Aadmi Party MLA from Gandhi Nagar(Delhi) Anil Bajpayi joins BJP in presence of Union Minister Vijay Goel

Aam Aadmi Party MLA from Gandhi Nagar Anil Bajpai Friday joined the BJP, making him the second lawmaker from the party to join the saffron fold in the last two months. The move comes at a time when AAP leaders, including Delhi Chief Minister Arvind Kejriwal, have accused the BJP of indulging in....

Support_Multimodal

(c)

Prime Minister Narendra Modi introduces council of ministers to the Lok Sabha.

Prime Minister Narendra Modi on Tuesday treated the Congress as his punching bag in the Lok Sabha during his reply to a discussion in Lok Sabha, rebutting criticism that his government had attempted to corner credit for India's development and belittled his predecessors. "I may be the only prime min......

Insufficient_Multimodal

(d)

Jharkhand: Six members of a family in Dhanbad have lost their lives due to COVID19. Civil surgeon, Dhanbad says, "2 more members of the same family have tested positive for COVID19, their condition is stable. We are doing their contact tracing."

Three inmates of prisons in Gautam Buddh Nagar and Ghaziabad have contracted Covid-19 and are undergoing treatment, officials said on Thursday. Of the three prisoners, two are inmates of Dasna jail in Ghaziabad and one of Luksar jail in Gautam Buddh Nagar, a senior official said. While one.....

Insufficient_Text

(e)

Swati Singh

Gosh - so much progress under Modiji! Women farmers are provided helicopter services to go to their fields for harvesting! 🤣

2:40 PM - 3 Apr 2019

1,615 Retweets 3,860 Likes

Many social media handles -- Facebook pages, users and Twitter users -- are sharing two photos of Hema Malini, one in which she's posing sitting in a helicopter and the other, posing as if she's lending a helping hand to female farmers, with a claim that the BJP leader is being "airdropped across wheat farms for photo-ops ....

Payal Mehta/પાયલ મેહતા/ पायल मेह...

Helicopter farmer @dreamgirlhema #NamumkinAbMumkinHai #ModiHaiTohMumkinHai Pic via: special arrangement

3:54 PM - 3 Apr 2019 from New Delhi, India

266 Retweets 546 Likes

Journalist Payal Mehta tweeted these photos on April 3 saying it was made available via "special arrangement". The same photos were shared on Facebook as well: TRUTHThese photos of Hema Malini that are being widely shared across social media platforms with

Refute

(f)

Figure 1: Examples for all the 5 categories

The objective of this shared task is to detect multimodal fake news, with each data point containing a true information source and another source whose validity has to be verified. The true/reliable source is called "document" and the other information source is called "claim". Both source and claim information sources are multimodal (they both have a corresponding image and text). The task is to check the validity of the multimodal claim with the document, ie, whether the claim entails the document. Essentially, given a textual claim, claim image, document and document image, the task is to classify the data point into one of the following five classes:

- `Support_Text`: the claim text is entailed but the document image and claim image are not similar.
- `Support_Multimodal`: both the claim text and image are similar to that of the document.
- `Insufficient_Text`: both text and images of the claim are neither similar nor refuted by the document, although it is possible that the text claim has common words with the document text.
- `Insufficient_Multimodal`: the claim text is neither similar nor refuted by the document but the claim image is similar to the document image.

- `Refute`: The images and/or text from the claim and document are contradictory i.e, the claim is false.

Figure 1 shows some examples of these classes. The task page is available on Codalab at https://competitions.codalab.org/competitions/35153.

3.1. Data

We use the dataset provided [10]. by For data creation, day-wises tweets from the twitter handles of two prominent news sources (Hindustan Times and ANI for India, ABC and CNN for US), were collected. From each tweet, text and image were extracted. Text similarity and image similarity was measured between each pair of tweets from both accounts, for a given day. If both text and image were similar for a tweet pair, they were classified as `Support_Multimodal`, and if text was similar but the images were not, then the tweet pair was classified as `Support_Text`. If both text and image were not similar for a tweet pair, they were classified as `Insufficient_Text`, and if the images were similar but the text was not, then the tweet pair was classified as `Insufficient_Multimodal`. For each of these four categories, one of the tweet text was replaced by the corresponding news article. For refute category, several fact-checking websites were scraped for fake news claims, the article refuting that claim, the fake news image and the image that proves that the news is fake.

The dataset has a total of 50,000 samples, and each of the 5 classes have equal sample size. The train-val-test split of the dataset is 70:15:15. For more details, please refer [10].

3.2. Baseline

For the purpose of the task, we provide the participants with the multi-modal baseline as given by [10]. The model uses both image and text. Image embeddings are obtained through a pretrained ResNet50 model. The text embeddings are obtained through a pre-trained SBERT model, 'paraphrase-MiniLM-L6-v2'. Cosine similarity is calculated for both image and text features, and are fed to a random forest classifier. The architecture can be visualised in Figure 2.

3.3. Evaluation

The task is essentially a multi-label classification problem. For determining the performance of a system, we use the average F1 score on the five categories: “Support_Text“, “Support_Multimodal“, “Insufficient_Text“, “Insufficient_Multimodal“ and “Refute“.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}, Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

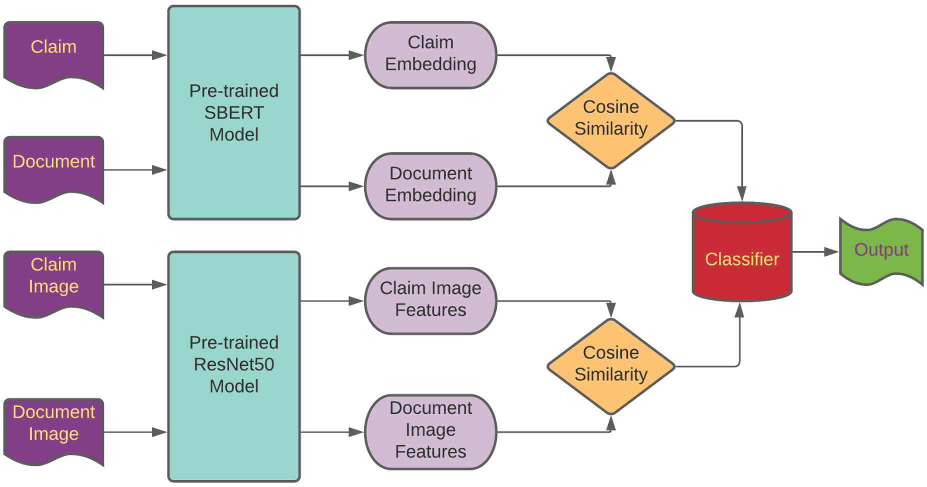


Figure 2: Multi-Modal Baseline Model which takes claim text, claim image as well as document text and document image as input.

For each category, F1 score is calculated. Since the number of samples for each category are equal, we determine the final score by simply averaging the F1 score over each category. The participants were asked to submit at most 3 runs on the test data, and the best one was chosen for the leaderboard.

4. Participating systems

4 teams initially participated in the shared task out of which 9 teams submitted test results. 8 teams submitted their system description papers. In this section, we provide an overview of the methods used by the teams.

Tyche [29] used BERT [30] to extract features from claim text and document text. EfficientNetB3 [31] is used to extract features from document image and claim image. These features are concatenated and given to fully connected layer followed by softmax classifier.

Yao [32] use pre-trained DeiT [33] and DeBERTa [34] to extract features from images and texts respectively. These features are fused in a co-attention model [35] having branches which fuses document image - claim image, document text - claim text, document image - document text, claim image - claim text. The outputs of each branch are aggregated and given as input to a softmax classifier for prediction.

Logically [36] train a decision tree classifier [37] which their image features and text features. The text features include text entailment predicted using BigBird [38], claim text length and document text length. The image features include claim and document source domain, claim and document image OCR length, and the image similarity score calculated using ResNet-50 [39].

Greeny [40] extract document and claim text features from RoBERTa [41] while extracting claim and document image features from ResNet-50 model [39]. These features are combined in a Gradient Booster [42] to make the final prediction.

Yet [43] employ pre-trained RoBERTa [41] to get claim text and document text features, which are later concatenated and fed to an MLP. Similarly, claim image and document image features are extracted using pre-trained VGG-16 [44], then concatenated and fed to an MLP. The output of the to MLP models is given to a classifier for final prediction.

Truthformers [45] utilize BERT [30] to extract claim text and document text features and Vision Transformer [46] to extract claim image and document image features. They use Conv1d for feature fusion. The claim text and document text features are passed through a text Conv1d layer. Similarly, the image features are passed through a image conv1d layer. After that they create document and claim vectors by concatenating respective text and image. These vectors are passed though a fully connected layer, the output of which is concatenated and passed to a classifier for prediction. Further, they use pseudo labeling [47] to improve the results of their model.

UofA-Truth [48] break down the task into two sub-tasks: text entailment and image entailment. For text entailment, they obtain claim text and document text embeddings by passing the text through sentence BERT [49]. These embeddings are concatenated along with their cosine similarity and fed to Fully Connected layers for classification. Similarly, they solve image entailment by using Xception [50] to generate image embeddings. The outputs of the two sub-tasks are consolidated in post-processing to get predictions on the original task.

GPTs [51] attempt to solve the task by considering only text and OCR text. They train multiple BERT based models (RoBERTa [41], DeBERTa [34], XLM-RoBERTa [52] and ALBERT [53]) and ensemble their predictions to generate final predictions. Further, they improve the model performance by adding prompt based learning [54] to the models, which helps filter out *refute* class more accurately.

5. Results

Rank	Team	Support_Text	Support_Multimodal	Insufficient_Text	Insufficient_Multimodal	Refute	Final
1	**Logically** [36]	**81.843%**	87.429%	84.437%	78.345%	99.899%	**76.819%**
2	Yet [43]	75.518%	89.38%	82.121%	80.81%	99.866%	75.591%
3	Truthformers [45]	77.65%	85.057%	79.421%	84.482%	98.819%	74.862%
4	UofA-Truth [48]	78.493%	**89.786%**	82.995%	75.981%	98.339%	74.807%
5	Yao [32]	68.881%	81.61%	84.836%	**88.309%**	**100.0%**	74.585%
6	Greeny [40]	74.947%	86.018%	80.382%	82.858%	99.125%	74.28%
7	GPTs [51]	71.575%	79.032%	75.363%	79.275%	**100.0%**	69.461%
8	Tyche [29]	75.0%	75.259%	**85.496%**	68.823%	99.159%	69.203%
9	MUM_NLP	64.803%	80.857%	69.848%	66.548%	93.465%	61.165%
-	**BASELINE**	82.675%	75.466%	74.424%	69.678%	42.354%	53.098%

Table 1
Top 9 teams for the Factify task. The teams are ranked by the overall weighted average F1 score (Final). We also report the category-wise F1 score for each team.

Table 1 shows the results of the top 9 teams for the Factify task. All of them made significant improvements over the baseline score. The winning team achieved a 76.819% Final score. All teams were able to achieve a high score in the Refute category. This could be because the refute data was collected from a different source than other classes. This shows the models were easily able to identify the difference between Refute and other categories. The Support_Text category score shows a lot of variances across teams and could not improve over the baseline score. Further, all the teams performed better on Support_Multimodal than on Support_Text. The leader-board shows that there is a scope of improvement on most of the subcategories. It should also be noted that no single team did better than other teams in all the categories.

6. Conclusion and Future Work

In this paper, we describe and summarize the Factify (multi-modal fact verification) shared task. We see that BERT-based models for text and Convolutional Neural Network or vision transformers for images are the most popular feature extractors used by the winners and many participants. Ensemble techniques are also quite popular. We saw some interesting methods which are worth exploring further. From the results of Factify task, we can conclude that it is a difficult task to create a model which performs well in all categories. The shared task reported in this paper aims to detect fake news, however, this problem is far from solved and requires further research attention. Future work could involve creating datasets for more languages and providing an explanation of why the post is fake. Another direction could be to provide the levels of fakeness instead of simple yes/no.

References

[1] J. Thorne, A. Vlachos, C. Christodoulopoulos, A. Mittal, Fever: a large-scale dataset for fact extraction and verification, arXiv preprint arXiv:1803.05355 (2018).

[2] A. Hanselowski, A. PVS, B. Schiller, F. Caspelherr, D. Chaudhuri, C. M. Meyer, I. Gurevych, A retrospective analysis of the fake news challenge stance-detection task, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 1859–1874. URL: https://aclanthology.org/C18-1158.

[3] W. Y. Wang, " liar, liar pants on fire": A new benchmark dataset for fake news detection, arXiv preprint arXiv:1705.00648 (2017).

[4] T. Mitra, E. Gilbert, Credbank: A large-scale social media corpus with associated credibility annotations, in: ICWSM, 2015.

[5] R. Mihalcea, C. Strapparava, The lie detector: Explorations in the automatic recognition of deceptive language, in: Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, ACLShort '09, Association for Computational Linguistics, USA, 2009, p. 309–312.

[6] I. Augenstein, C. Lioma, D. Wang, L. Chaves Lima, C. Hansen, C. Hansen, J. Grue Simonsen, Multifc: A real-world multi-domain dataset for evidence-based fact checking of claims, in: EMNLP, Association for Computational Linguistics, 2019.

[7] K. Nakamura, S. Levy, W. Y. Wang, r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection, arXiv preprint arXiv:1911.03854 (2019).

[8] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media, 2019. arXiv:1809.01286.

[9] J. C. S. Reis, P. de Freitas Melo, K. Garimella, J. M. Almeida, D. Eckles, F. Benevenuto, A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections, 2020. arXiv:2005.02443.

[10] S. Mishra, S. Suryavardan, A. Bhaskar, P. Chopra, A. Reganti, P. Patwa, A. Das, T. Chakraborty, A. Sheth, A. Ekbal, et al., Factify: A multi-modal fact verification dataset, in: De-Factify workshop at AAAI, 2022.

[11] G. Pennycook, D. G. Rand, The psychology of fake news, Trends in Cognitive Sciences 25 (2021) 388–402. URL: https://www.sciencedirect.com/science/article/pii/S1364661321000516. doi:https://doi.org/10.1016/j.tics.2021.02.007.

[12] A. R. Ahmad, H. R. Murad, The impact of social media on panic during the covid-19 pandemic in iraqi kurdistan: Online questionnaire study, Journal of Medical Internet Research 22 (2020).

[13] S. Loomba, A. Figueiredo, S. Piatek, K. de Graaf, H. Larson, Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa, Nature Human Behaviour 5 (2021) 1–12. doi:10.1038/s41562-021-01056-1.

[14] C. Martel, G. Pennycook, Reliance on emotion promotes belief in fake news, Cognitive research: principles and implications 5 (2020) 47. doi:10.1186/s41235-020-00252-3.

[15] S. Talwar, A. Dhir, D. Singh, G. S. Virk, J. Salo, Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis, Journal of Retailing and Consumer Services 57 (2020) 102197. URL: https://www.sciencedirect.com/

science/article/pii/S0969698920306433. doi:https://doi.org/10.1016/j.jretconser.2020.102197.

[16] P. Patwa, M. Bhardwaj, V. Guptha, G. Kumari, S. Sharma, S. PYKL, A. Das, A. Ekbal, S. Akhtar, T. Chakraborty, Overview of constraint 2021 shared tasks: Detecting english covid-19 fake news and hindi hostile posts, in: Proceedings of the First Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT), Springer, 2021.

[17] J. Bevendorff, B. Ghanem, A. Giachanou, M. Kestemont, E. Manjavacas, I. Markov, M. Mayerl, M. Potthast, F. Rangel Pardo, P. Rosso, G. Specht, E. Stamatatos, B. Stein, M. Wiegmann, E. Zangerle, Overview of PAN 2020: Authorship Verification, Celebrity Profiling, Profiling Fake News Spreaders on Twitter, and Style Change Detection, 2020, pp. 372–383. doi:10.1007/978-3-030-58219-7_25.

[18] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, C. C. Ferrer, The deepfake detection challenge (dfdc) dataset, 2020. arXiv:2006.07397.

[19] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, T. Chakraborty, Fighting an infodemic: Covid-19 fake news dataset, in: Combating Online Hostile Posts in Regional Languages during Emergency Situation (CONSTRAINT) 2021, Springer, 2021, p. 21–29. URL: http://dx.doi.org/10.1007/978-3-030-73696-5_3. doi:10.1007/978-3-030-73696-5_3.

[20] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, S. Satoh, Spotfake: A multi-modal framework for fake news detection, in: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), 2019, pp. 39–47. doi:10.1109/BigMM.2019.00-44.

[21] J. Ma, W. Gao, K.-F. Wong, Rumor detection on Twitter with tree-structured recursive neural networks, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 1980–1989. URL: https://aclanthology.org/P18-1184. doi:10.18653/v1/P18-1184.

[22] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, J. Gao, Eann: Event adversarial neural networks for multi-modal fake news detection, 2018, pp. 849–857. doi:10.1145/3219819.3219903.

[23] D. Khattar, J. Singh, M. Gupta, V. Varma, Mvae: Multimodal variational autoencoder for fake news detection, 2019, pp. 2915–2921. doi:10.1145/3308558.3313552.

[24] Z. Liu, C. Xiong, M. Sun, Z. Liu, Fine-grained fact verification with kernel graph attention network, arXiv preprint arXiv:1910.09796 (2019).

[25] P. Qi, J. Cao, T. Yang, J. Guo, J. Li, Exploiting multi-domain visual information for fake news detection, 2019. arXiv:1908.04472.

[26] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, 2017. arXiv:1708.07104.

[27] Z. Guo, M. Schlichtkrull, A. Vlachos, A survey on automated fact-checking, Transactions of the Association for Computational Linguistics 10 (2022) 178–206.

[28] X. Zhou, R. Zafarani, A survey of fake news, ACM Computing Surveys 53 (2020) 1–40. URL: http://dx.doi.org/10.1145/3395046. doi:10.1145/3395046.

[29] A. Raj, N. Hulke, A. A. Saifee, B. R. Siva, Tyche at factify 2022: Fusion networks for multi-modal fact-checking, in: Proceedings of De-Factify: Workshop on Multimodal Fact

Checking and Hate Speech Detection, CEUR, 2022.

[30] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[31] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR, 2019, pp. 6105–6114.

[32] W.-Y. Wang, W.-C. Peng, Team Yao at factify 2022: Utilizing pre-trained models and co-attention networks for multi-modal fact verification, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[33] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jegou, Training data-efficient image transformers &; distillation through attention, in: M. Meila, T. Zhang (Eds.), Proceedings of the 38th International Conference on Machine Learning, volume 139 of *Proceedings of Machine Learning Research*, PMLR, 2021, pp. 10347–10357. URL: https://proceedings.mlr.press/v139/touvron21a.html.

[34] P. He, X. Liu, J. Gao, W. Chen, DEBERTA: Decoding-enhanced bert with disentangled attention, in: International Conference on Learning Representations, 2021. URL: https://openreview.net/forum?id=XPZIaotutsD.

[35] Y. Wu, P. Zhan, Y. Zhang, L. Wang, Z. Xu, Multimodal fusion with co-attention networks for fake news detection, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 2560–2569.

[36] J. Gao, H.-F. Hoffmann, S. Oikonomou, D. Kiskovski, A. Bandhakavi, Logically at the factify 2022: Multimodal fact verfication, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[38] M. Zaheer, G. Guruganesh, K. A. Dubey, J. Ainslie, C. Alberti, S. Ontanon, P. Pham, A. Ravula, Q. Wang, L. Yang, et al., Big bird: Transformers for longer sequences, Advances in Neural Information Processing Systems 33 (2020) 17283–17297.

[39] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[40] W. Bai, Greeny at Factify 2022: Ensemble model with optimized roberta for multi-modal fact verification, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[41] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, RoBERTa: A robustly optimized bert pretraining approach (2019).

[42] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, Advances in neural information processing systems 30 (2017) 3146–3154.

[43] Y. Zhuang, Y. Zhang, Yet at Factify 2022 : Unimodal and bimodal roberta-based models for fact checking, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[44] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[45] C. B. S. N. V, P. Potluri, R. Vijjali, Truthformers at Factify 2022 : Evidence aware transformer based model for multimodal fact checking, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[46] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, ICLR (2021).

[47] D.-H. Lee, Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks, ICML 2013 Workshop : Challenges in Representation Learning (WREPL) (2013).

[48] A. Dhankar, O. Zaiane, F. Bolduc, UofA-Truth at Factify 2022 : A simple approach to multi-modal fact-checking, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[49] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019. URL: https://arxiv.org/abs/1908.10084.

[50] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.

[51] S. Aggarwal, P. Sahu, T. Gupta, G. Das, GPTs at Factify 2022: Prompt aided fact-verification, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[52] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, arXiv preprint arXiv:1911.02116 (2019).

[53] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: A lite bert for self-supervised learning of language representations, arXiv preprint arXiv:1909.11942 (2019).

[54] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.