

Detection of Persuasion Techniques in Memes

Premkanna Gopalakrishnan¹, Gokul Ajith V², Chandrabose Aravindan³ and R. Subhashini⁴

^{1,3,4}Sri Sivasubramaniya Nadar College of Engineering, Department of Information Technology Chennai, India
Email: premkannajg@gmail.com, aravindanc@ssn.edu.in, subhaagopi@gmail.com

²SRM Institute of Science and Technology, Department of CSE, Ramapuram, Chennai, India
Email: gokul2012053@ssn.edu.in

Abstract— In today's digital sphere, memes wield incredible influence, especially when used in deceptive campaigns. Their unique blend of visuals and text allows for rapid dissemination on social platforms, making them a potent tool for spreading disinformation. Understanding the tactics employed within these memes ranging from subtle rhetorical strategies to emotional appeals is crucial in combating online deception. This paper delves into the task of identifying and categorizing these persuasion techniques within memes, employing a comprehensive approach that analyzes both textual and visual elements.

Keywords— NLP – Natural Language Processing • BERT – Bidirectional Encoder Representations for Transformer • RoBERTa – Robustly Optimized BERT Pre-training Approach • DeBERTa – Decoding-enhanced BERT with Disentangled Attention

I. INTRODUCTION

It's fascinating how social networks have evolved to become pivotal in shaping our society. Nowadays, they serve as conduits for sharing ideas, thoughts, and political beliefs, predominantly through platforms like Twitter, Facebook, and Instagram. While these platforms efficiently disseminate information, it's increasingly challenging to discern whether the content shared is free from subtle manipulations that might impact people's judgment. Among the myriad content types inhabiting social networks, memes wield considerable influence. These concise yet impactful units of information propagate cultural ideas, symbols, or practices primarily in the form of images coupled with text overlays.

Memes are engineered for rapid dissemination, making them a favored tool in online disinformation campaigns. These campaigns leverage various rhetorical and psychological strategies—such as causal oversimplification, name-calling, or smear tactics to sway users' perceptions.

Detecting these memes and unraveling the disinformation techniques they employ pose a complex challenge for managing social networks effectively.

In recent years, the advancements in machine learning and deep learning have marked significant milestones in extracting and interpreting multimedia data automatically. These breakthroughs play a crucial role in analyzing multimedia content at scale within social networks.

Memes have emerged as a prevalent medium for disseminating information, often intertwined with online disinformation campaigns.

These campaigns utilize various rhetorical and psychological techniques to influence and manipulate users on social media platforms. The Semeval task 4 2024 shared task is specifically centered around developing models capable of identifying these persuasive techniques within both the textual and visual content of memes.

IV. LITERATURE SURVEY

LIIR at SemEval-2021 task 6: Detection of Persuasion Techniques In Texts and Images using CLIP features.: Erfan Ghadery , Damien Sileo and Marie-Francine Moens. This system performs competitively with other systems even though we used a simple architecture with no ensemble, by leveraging non- supervised learning. We believe that further work on zero-shot learning would be a valuable way to improve propaganda detection techniques for the least frequent labels. Detecting and Understanding Harmful Memes: A Survey.: Shivam Sharma^{1,2}, Firoj Alam³, Md. Shad Akhtar¹, Dimitar. Surveying contemporary intelligent technologies in harmful meme detection, our work introduces a novel meme typology, consolidating insights from recent literature. Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. 2018. Bottom-up and top-down attention for image captioning and visual question answering. In Proc. of the IEEE conference on computer vision and pattern recognition, pages 6077– 6086. : Innovative approach integrating bottom-up and top- down attention mechanisms for improved image captioning and visual question answering. Maram Hasanain, Ahmed El-Shangiti, Rabindra Nath Nandi, Preslav Nakov, and Firoj Alam. 2023b. QCRI at SemEval-2023 task 3: News genre, framing and persuasion techniques detection using multilingual models. In Proceedings of the 17th International: Addresses multilingual models for detecting news genre, framing, and persuasion techniques, likely highlighting advancements in language processing for these tasks. Maram Hasanain, Firoj Alam, Hamdy Mubarak, Samir Abdaljalil, Wajdi Zaghoulani, Preslav Nakov, Giovanni Da San Martino, and Abed Alhakim Freihat. 2023a. ArRIval Shared Task: Persuasion Techniques and Disinformation Detection in Arabic Text. Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023), Singapore. Association for Computational Linguistics.: Aims to detect persuasion techniques and disinformation in Arabic text, presented at the ArabicNLP 2023 conference. Nicolas Carrion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End- to-end object detection with transformers. In the European Conference on Computer Vision, pages 213– 229. Springer: Aims to detect persuasion techniques and disinformation in Arabic text, presented at the ArabicNLP 2023 conference.: Discusses object detection using transformer-based models, potentially introducing an end-to-end approach for this task. Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. 2019. Show, control and tell: A framework for generating controllable and grounded captions. In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pages 8307– 8316.: Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT 2019, pages 4171–4186. Association for Computational Linguistics: Introduces BERT, a language representation model that revolutionized natural language understanding tasks through pre-training on large text corpora. Abujaber Dia, Qarqaz Ahmed, and Abdullah Malak A. 2021. LeCun at SemEval-2021 Task 6: Detecting persuasion techniques in text using ensembled pretrained transformers and data augmentation. In Proceedings of the International Workshop on Semantic Evaluation, SemEval '21, Bangkok, Thailand

V. THE DATASET

We amassed a collection of English memes from social media via looking through various public businesses. These agencies blanketed topics starting from politics and vaccines to COVID-19, gender equality, and the Russo-Ukrainian conflict. To ensure consistency, we described a meme as a picture with a quick snippet of text masking it. We filtered out memes that failed to match this definition, along with cartoon-fashion memes or those with minimal textual content. Subsequent, we meticulously categorized every meme the usage of 22 specific persuasion techniques.

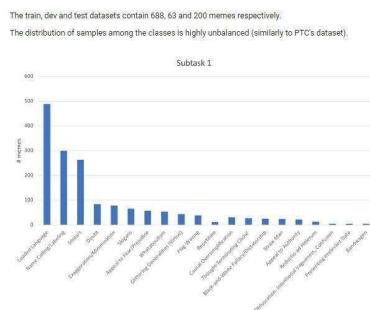


Fig:1 Task 1 dataset

Our annotation process worried about dissecting the text of each meme after analyzing the meme as an entire. To keep information integrity, we divided our dataset into separate sets for education, validation, development, and testing. no matter the potential for longer textual content, we determined that most people of memes contained minimal textual content, with an average of just 1. sixty eight sentences in line with meme. even as some memes had up to 13 sentences, they were the exception in place of the norm.

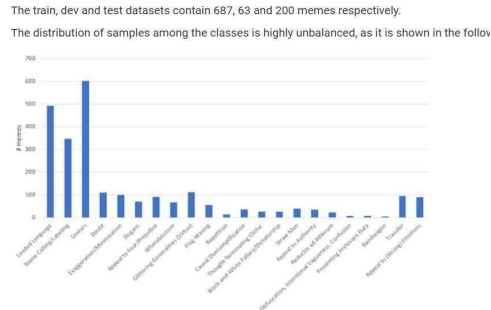


Fig:2 Task 2 dataset

VI. PROPOSED METHODOLOGY

Our proposed methodology covers many important steps to solving the task at hand. First, during the data preparation phase, we collect important data, including text and related documents. We then carefully preprocessed the data, retained missing values, and labeled and coded the text. We then split the data into training and testing subsets to facilitate model training and evaluation.

He is known for his achievements in natural language understanding studies. We carefully tailor the selection of training materials to suit the characteristics of our business. To solve the class ambiguity problem that often arises in real-world data, we carefully select the training and weight loss during training, using the focus loss function together with AdamW's optimizer configuration. We rigorously evaluate the training model using a number of performance metrics such as accuracy, precision, recall, and F1 score. Additionally, we used cross-validation to verify the generality and robustness of the model across different data sets. Hyperparameter tuning further optimizes model performance by focusing on parameters such as learning rate and batch size to achieve the best results.

We have prepared a collection of various titles paired with matching images. First, data files were created and images were resized and normalized. We leverage the latest models to generate embeddings from two algorithms: DeBERTa for text and the faster R-CNN with ResNet50 backbone for images. These embeddings are combined to form a unified representation that allows the model to understand the relationship between text and images. Classification is then done using a transformation tool such as DeBERTa and then a layer is applied for final classification.

We seamlessly integrate with production systems via API or Web services to ensure delivery and efficient operation. The training and testing process is defined by PyTorch Lightning's Trainer class; where “trainer.fit(model, dm)” handles the training model and “trainer.test(model, dm)” makes the evaluation model to provide the evaluation model.

VII. FLOW CHARTS/ ARCHITECTURE

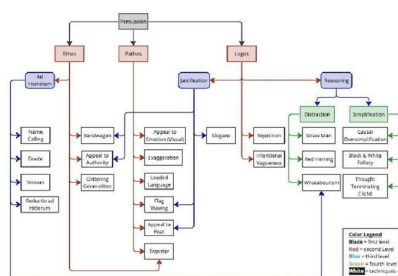


Fig 3. Persuasion techniques

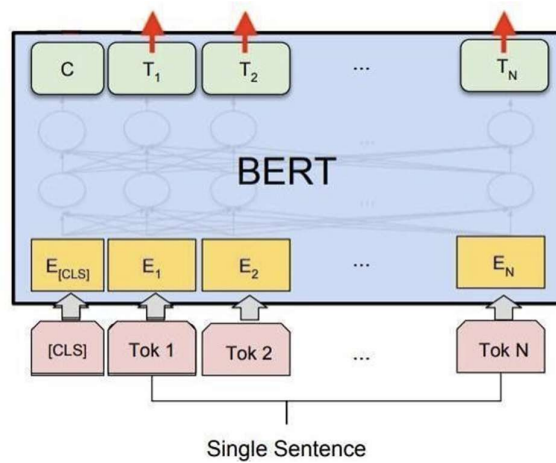


Fig 4. BERT model structure

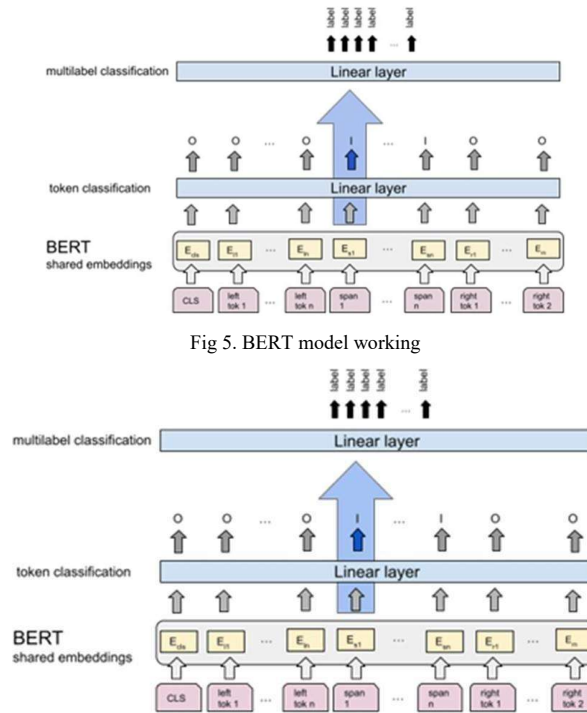


Fig 5. BERT model working

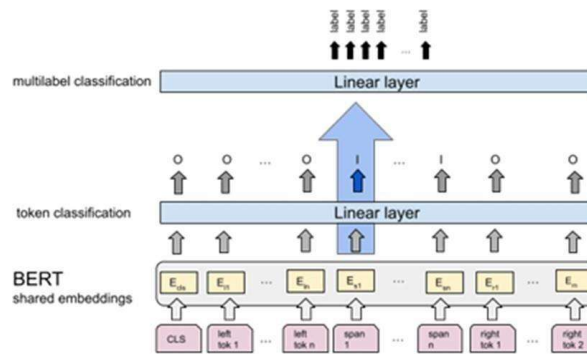


Fig 6. Work sequence for task1

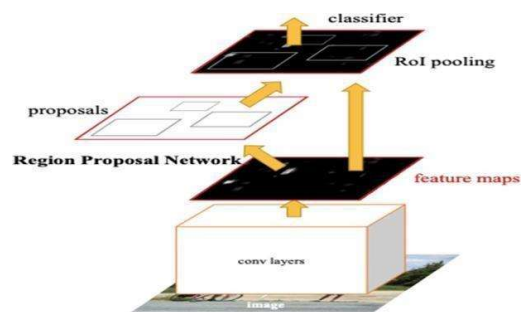


Fig 7. Architecture of DeBERTa + Resnet50 Embedding

VIII. EVALUATION METRICS

Subtask1 is a multilabel classification problem. We used $-F1$ as the official evaluation measure. We also computed precision and recall. Citepaper for metrics subtask 2 is a multimodal, multilabel classification problem. We used $-F1$ as the official evaluation measure as the official evaluation measure.

TABLE I. TASK1 METRICS AND SCORES

	Name	Type	Params
0	loss	FocalLoss	0
1	acc	Accuracy	0
2	transformer	PTCTransformer	108 M
3	classifier	Sequential	75.8 K

Fit end: Evaluation....
F1-micro: 0.5380710659898477
F1-macro: 0.2981065759637188

TABLE II. DEBERTA PERFORMANCE SCORE

	Name	Type	Params
0	loss	BCEWithLogitsLoss	0
1	acc	Accuracy	0
2	fusion_encoder	FusionEmbeddings	181 M
3	transformer	DebertaModel	138 M
4	classifier	Sequential	50.6 K

F1-micro: 0.5113268608414239
F1-macro: 0.20582862113962594

IX. CONCLUSION

The proposed project represents a significant step forward in the fight against online disinformation, providing valuable tools for identifying and understanding the persuasive tactics employed in memes. Through ongoing research and collaboration, we remain committed to advancing the field of misinformation detection and fostering a more informed and resilient digital society

REFERENCES

- [1] Afridi et al., 2021] Tariq Habib Afridi, Aftab Alam, Muhammad Numan Khan, Jawad Khan, and Young Koo Lee. A multimodal memes classification: A survey and open research issues. In SCA, pages 1451–1466, 2021
- [2] Erfan Ghadery and Marie-Francine Moens. 2020. Liir at semeval-2020 task 12: A cross-lingual augmentation approach for multilingual offensive language identification. arXiv preprint arXiv:2005.03695.
- [3] Erfan Ghadery, Sajad Movahedi, Heshaam Faili, and Azadeh Shakery. 2018. An unsupervised approach for aspect category detection using soft cosine similarity measure. arXiv preprint arXiv:1812.03361.
- [4] Marco Guerini, Carlo Strapparava, and Oliviero Stock. 2008. Resources for persuasion. Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08), Marrakech, Morocco. European Language Resources Association (ELRA).
- [5] Ivan Habernal and Iryna Gurevych. 2016. What makes a convincing argument? empirical analysis and detecting attributes of convincingness in web argumentation. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1214–1223.
- [6] Gasparini et al., 2021] Francesca Gasparini, Giulia Rizzi, Aurora Saibene, and Elisabetta Fersini. Benchmark dataset of memes with text transcriptions for automatic detection of multi-modal misogynistic content. arXiv:2106.08409, 2021.
- [7] Weijie Liu, Peng Zhou, Zhiruo Wang, Zhe Zhao, Haotang Deng, and Qi Ju. 2020. FastBERT: a self distilling BERT with adaptive inference time. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL ’20, pages 6035–6044, Online.
- [8] Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee. 2019. ViLBERT: Pre Training task-agnostic visio linguistic representations for vision-and-language tasks. In Proceedings of the Conference on Neural Information Processing Systems, NeurIPS ’19, pages 13–23, Vancouver, Canada.

- [9] Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020. A survey on computational propaganda detection. In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI-PRICAI '20*, pages 4826–4832.
- [10] Konrad Kaczyński and Piotr Przybyła. 2021. HOMADOS at SemEval-2021 Task 6: Multi-task learning for propaganda detection. In *Proceedings of the International Workshop on Semantic Evaluation, SemEval '21*, Bangkok, Thailand.
- [11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *Proceedings of the 1st International Conference on Learning Representations, Workshop Track, ICLR '13*, Scottsdale, Arizona, USA.