Findings of Memotion 2: Sentiment and Emotion Analysis of Memes

Parth Patwa[1], Sathyanarayanan Ramamoorthy[2], Nethra Gunti[2], Shreyash Mishra[2], S Suryavardan[2], Aishwarya Reganti[3], Amitava Das[4,5], Tanmoy Chakraborty[6], Amit Sheth[5], Asif Ekbal[7] and Chaitanya Ahuja[8]

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

Abstract

Memes are an important part of the Internet culture and their popularity has increased in the recent years. Memes can be used to express humor, opinions or to even spread hate and misinformation. Hence, it is of research interest to analyze them. In this paper, we describe the Memotion 2 shared task, which is organized as a part of the De-Factify workshop at AAAI'22. The shared task includes study of memes in three sub-tasks – Task A: sentiment analysis, Task B: emotion analysis, Task C: emotion intensity detection. A total of 44 teams participated in the Memotion 2.0 shared task, and of them, 8 teams submitted their predictions on test set for Tasks A an B, and 7 teams for Task C. Use of BERT-like models was a popular choice to extract text features among the participants. Models like ResNet50, VGG-16, EfficientNet were used by the participants to extract text features. Most of the systems combine the modalities (text,image) in a late fusion. The best F1 scores achieved for the Tasks A, B and C are 0.53, 0.82 and 0.55, respectively.

Keywords

Memes, Sentiment Analysis, Dataset, Multimodality

1. Introduction

The word 'meme' was first used by Richard Dawkins in his 1976 book, "The Selfish Gene" [1], calling it as cultural units that replicate, mutate and evolve. Modern memes are an extension of the original idea of a meme, but the mode of spread is through online platforms. Memes have become a very popular mode of broadcast and communication over social media platforms these

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); sathyanarayanan.r18@iiits.in (S. Ramamoorthy); nethra.g18@iiits.in (N. Gunti); shreyash.m19@iiits.in (S. Mishra); suryavardan.s19@iiits.in (S. Suryavardan); amitava.das2@wipro.com (A. Das)

days. The usage of simple language to convey a message resonates with the general public, and the reach for such posts is humongous. There is a psychological side of this phenomenon of rapid sharing of memes. [2] examined responses of individuals towards Internet videos, memes and found that strong affective responses to a video reported greater intent to spread it across social platforms. Usually the content of a meme is derived from day-to-day activities, college life, work environment, food, relationships, etc. Therefore, people develop more affinity towards such posts that explains the reason behind the viral reach of memes across countries in a shorter span of time.

The study of evolution of memes over the last ten years helps us to also understand the changes in online culture. Initially, memes served as an expression of an individual's take on a subject with some humor and little sarcasm. Because of the freedom that one enjoys while creating memes, they were then used to vent one's feelings on socio-political issues. Currently, we are in a stage where the memes are being used by social media users to share their opinions on any topic, which further helps them connect with millions of people all over the globe. Memes are succinct and explicit in their messages. A flip side of this powerful medium is that it is being misused to spread hatred in the community. The work by Moody and Church [3] analyzes the role that Facebook meme pages and trolls had in the 2016 US Presidential Elections. There has been an increase in the number of online abuses, especially on the oppressed and weaker sections of the society. [4] shows that many Internet memes featuring fake news are specifically directed with political agendas by agencies. An article [5] unpacking the vulgarity of Internet memes targeting aboriginality in relation to skin colour and other racist stereotypes, is a reminder to the research community that identifying and preventing toxic news present in social media is necessary. Recently, a few models have been proposed to detect harmful memes and their targets [6, 7]

This paper presents the findings of the shared task Memotion 2, which was organized as part of the workshop *"Defactify - A workshop on Multimodal Fact-Checking and Hate Speech Detection"* at AAAI 2022. Our work is an attempt to leverage the information present in Internet memes and encourage research teams to develop robust computational methods to classify the sentiment, emotion and emotion intensity of multi-modal posts (memes) accurately.

2. Related Work

Mining and understanding social media content has become a very significant task in recent times. There exists an abundance of data in different forms and tenor. Extracting this information can help examine diffusion, recommendations, analyze behaviour, etc. This ever-growing flow of diverse content has attracted researchers to several applications using online data, one of of which is sentiment and emotion analysis. The task aims to identify and quantify subjective information from given data. Most of the studies in this area focus on collecting relevant public data and applying it to binary, polarity or scale based classification task [8]. Some of the notable contributions include textual and multi-modal datasets [9] [10] [11], workshops [12], and modelling approaches [13]. The second aspect is hate speech detection. Automation of such a task is challenging because of the vast variety of content and the blurry line between hate and free speech. Research towards hate speech detection presents workshops [14] [15]

Figure 1: In the dataset, this meme is labelled as having neutral sentiment and emotions are funny, little sarcastic, slightly offensive and not motivational. For **Task A**, the predicted sentiment should be **neutral**. For **Task B**, the emotions predicted will be **funny, sarcastic, offensive and non motivational**. In **Task C**, we expect the system to predict the intensity of each emotion i.e for this meme, the predictions should be **funny, little sarcastic, slightly offensive and not motivational**. Refer to [31] for more details on the dataset.

and tasks with multi-lingual [16] [17] [18] and multi-modal [19] [20] data [21] [22]

An extensive amount of multi-modal social media data is in the form of memes. The attention to understanding and extracting sentiment, emotion and profanity from memes is growing. Meme analysis highlights the importance of considering both visual and textual cues to understand the context, offensiveness, humour, etc. The previous iteration of our task – Memotion 1.0 [23] provided an annotated dataset with labels capturing humour, sarcasm and hate speech. Other significant work in this area is from the Hateful memes challenge by Facebook [24], the MultiOFF dataset [25] and other such tasks [26, 27]. These tasks have brought attention to analysis of memes and (CNN, BERT, CLIP etc. based) multi-modal modelling approaches [6, 28, 29, 30].

3. Task Details

The idea behind the shared task is to facilitate the research community to analyze memes across multiple dimensions.

3.1. Tasks

We conduct and evaluate participants in three tasks:

- **Task A: Sentiment Analysis** - The task is to classify a given meme's sentiment as positive, negative or neutral.

- **Task B: Emotion Classification** - Identifying particular emotions associated with a given meme is the motive of this task. The system/model should indicate if the meme is humorous, sarcastic, offensive and motivational. A meme can belong to more than one category.
- **Task C: Scales/Intensity of Emotion Classes** - As humans, we express the same emotion in different levels of intensity . Hence, the third task is to quantify the extent to which a particular emotion is being expressed by a given meme. Different intensities of each emotion are:
  - **Humour**: Not funny, funny, very funny and hilarious
  - **Sarcasm**: Not Sarcastic, little sarcastic, very sarcastic and extremely sarcastic
  - **Offensive**: Not offensive, slightly offensive, very offensive and hateful offensive
  - **Motivation**: Not motivational, motivational

Tasks A, B and C are explained using the meme in Figure 1.

3.2. Dataset

The tasks were conducted as a part of Memotion 2 [31]. The memes were collected from various public platforms like Reddit, Instagram, etc. and annotated with the help of Amazon Mechanical Turk workers. The dataset consists of 10,000 meme images divided into a train-val-test split with 7000-1500-1500. Each meme is annotated for its overall sentiment, emotion and scale of each emotion. Images also have their corresponding OCR text extracted with the help of Google Vision APIs. For a detailed data description, please refer to [31].

3.3. Evaluation

The challenge involves three tasks, with Task A being a multi-label classification problem of identifying a sentiment (positive, neutral or negative) for a meme. Tasks B and C both are multi-label classification problems of emotion detection. Scoring is done for each task separately, and separate leaderboards are generated. For each task, we use weighted average F1 score to measure the performance of a model. The participants had access to only train and validation set. They were asked to submit a maximum of 3 submissions on the test set for each task, the best of which was selected as part of leaderboard.

3.4. Baselines

The baseline models for the tasks were created by keeping in mind the multi-modal nature of the dataset. BERT was used to extract text features from the OCR text for each meme and ResNet-50 for image features from the meme image. The features were then concatenated and passed on to a classification head to predict labels for each task For more details about the baseline, please refer [31].

Rank.	Team	F1 - Scores
1	**BLUE** [46]	**0.5318**
2	BROWALLIA [38]	0.5255
3	Yet [42]	0.5088
4	Little Flower [45]	0.5081
5	Greeny	0.5037
6	Amazon PARS [51]	0.5025
7	HCILab [32]	0.4995
8	weipengfei	0.4887
9	BASELINE	0.4340

Table 1
Leaderboard of teams on Task A: Sentiment Classification. All the teams cross the baseline.

4. Participating systems

Total 44 teams participated in the shared task, out of which 8 teams submitted their results for Tasks A and B, and 7 teams submitted papers for Task C. We received 6 system description papers. In this section, we provide a summary of the methods that the teams used.

HCIlab [32] used EfficientNet-v2 [33] for learning image embeddings and RoBERTa [34] for learning text embeddings. These embeddings were fused using a multihop attention mechanism [35], which as followed by a fully connected layer and a classifier. They also use auto augmentation [36] and CCA [37] to improve performance of their system.

BROWALLIA [38] used ResNet50 [39] and LSTM [40] to extract image and text embeddings, respectively. These embeddings are concatenated and given to a classifier. Further, they use offline-gradient-blending [41] to decrease overfitting. In this, they calculate the Overfiting-to-Generalization ratio and use it to weigh the loss function.

Yet [42] used VGG-16 [43] to extract image features and GloVe [44] followed by LSTM [40] to extract text features. These features are fused using fully-connected layers.

Little flower [45] used VGG-16 [43] followed by multi-head attention and dense layer along with residual connections to extract image features. For extracting text features, they used BiLSTM [40] followed by attention mechanism and fully-connected layers along with residual connection. The text and image features are concatenated in a late fusion. To get the final prediction, they used an ensemble method. Further, they used a weighted loss function to account for class imbalance.

BLUE [46] used a 3-branch network, where the branches use EfficientNetV4 [47], CLIP [48] and sentence transformer [49] respectively, for feature extraction. These features are given to a multi-task transformer encoder which makes the prediction. They trained the models using CORAL [50] loss function to predict the intensity of emotions.

Amazon PARS [51] used VisualBERT [52] to extract image features and BERT [53] to extract text features. These features are fed to a transformer. The transformer is trained in a two stage [54] multi-task manner, where the predictions of task B are fed to predict on Task C.

5. Results

Table 1 shows the leaderboard for Task A. All the participating teams managed to cross the baseline score for this task, with the relative increment between 12.6% and 22.5%. A top score of 0.5318 is achieved by BLUE [46] at an increment of 22.53% from the baseline score of 0.4340. The second top team, BROWALLI [38], is not far behind with a score of 0.5255, which is 21.08% increment from the baseline score. The remaining teams perform comparably with a small difference in their scores.

Rank.	Team	F1 - Scores				
		H	S	O	M	Overall
1	Little Flower [45]	**0.9384**	**0.819**	**0.5540**	**0.9800**	**0.8229**
2	BLUE [46]	**0.9384**	0.8183	0.4873	0.9797	0.8059
3	BROWALLIA [38]	0.9086	0.6705	0.5089	**0.9800**	0.7670
4	Amazon PARS [51]	0.9173	0.6282	0.5321	0.9658	0.7609
5	HCILab [32]	0.9124	0.5484	0.5247	**0.9800**	0.7414
6	BASELINE	0.7944	0.6575	0.5346	0.9575	0.7358
7	weipengfei	**0.9384**	0.3623	0.4862	0.9790	0.6915
8	Yet [42]	**0.9384**	0.0386	0.4853	**0.9800**	0.6106
9	Greeny	**0.9384**	0.0386	0.4853	**0.9800**	0.6106

Table 2
Leaderboard of teams on Task B: Emotion Classification {H:Humor, S:Sarcasm, O:Offense, M:Motivation}. The teams are ranked by their average F1 scores (overall) across all the four emotions. Motivation emotion is the easiest to detect while offense is the most difficult to detect.

Table 2 shows the leaderboard for Task B. Five teams managed to cross the overall baseline score, whereas three teams could not cross the baseline. The maximum scores for Humor, Sarcasm, Offense, and Motivation are 0.9384, 0.8190, 0.5540, 0.9800, respectively, which are at an increment of 14.41%, 16.14%, 1.94%, 2.25% from the baseline scores of each emotion, respectively. The overall top score of 0.8229, with an increment of 8.71% from baseline score of 0.7358, is achieved by Little Flower [45]. We can see that the 'Motivation' class is the easiest class to detect. Humor is also easy to detect, possibly because most of the memes are meant to be funny. Sarcasm is the hardest to detect and its scores vary considerably, whereas for other classes, the teams' scores are closer to each other.

Table 3 shows the leaderboard for Task C. Four teams cross the overall baseline score for this task, and four teams are below the baseline. The maximum scores for Humor, Sarcasm, Offense, and Motivation are 0.4611, 0.3083, 0.5275, 0.9800, respectively which are at an increment of 37.69%, 21.74%, 9.94%, 2.349% from the baseline scores of each emotion, respectively. The overall top score of 0.5564, with an increment of 9% from baseline score of 0.5105, is achieved by Amazon PARS [51]. The performance on 'Motivation' is much higher than on other emotions because motivation has only 2 intensities whereas other emotions have 4 intensities. All the teams perform poorly when detecting the intensity of sarcasm, which shows that most neural models fail to understand sarcasm. The best overall score is 0.554, which shows that there is a

Rank.	Team	F1 - Scores				
		H	S	O	M	Overall
1	Amazon PARS [51]	0.4598	0.2979	0.5021	0.9658	**0.5564**
2	BROWALLIA [38]	0.4508	0.2230	**0.5275**	**0.9800**	0.5453
3	BLUE [46]	0.4036	**0.3083**	0.4850	**0.9800**	0.5443
4	HCILab [32]	0.4212	0.2109	0.5144	0.9740	0.5301
5	BASELINE	0.3349	0.2533	0.4799	0.9575	0.5105
6	Yet [42]	0.4552	0.1194	0.4853	**0.9800**	0.5100
7	weipengfei	**0.4611**	0.0869	0.4862	0.9790	0.5033
8	Greeny	0.4435	0.0271	0.4853	**0.9800**	0.4840

Table 3
Leaderboard of teams on Task C: detection of intensity of emotion {H:Humor, S:Sarcasm, O:Offense, M:Motivation}. The teams are ranked by average F1 scores (overall) across all the four emotions. Intensity of sarcasm is by far the most difficult to detect. Motivation has only two intensities as opposed to four intensity levels for other emotions.

lot of scope of improvement.

Four teams, namely, BLUE [46], HCIlab [32], Amazon PARS [51], and BROWALLIA [38]; crossed the baseline scores for all the three tasks. Since Task B is a multi-task binary classification task, its results are better than those of Task A, which is a multi-class classification task. Results are better on Task B than Task C, because Task C is more fine grained.

6. Conclusion and Future Work

In this paper, we report the findings of the Memotion 2. We see that all the teams used deep learning based architectures, and most of the teams use BERT based models to extract language features. On the other hand, we see more variety in models used to extract image features (VGG, ResNet, EfficientNet, etc). Further, most systems use late fusion to combine image and text modalities. The best results on Task A (Sentiment Analysis), Task B (Emotion detection), and Task C (Emotion Intensity Detection) are 0.53, 0.82, 0.55, respectively, which shows that there is a large scope of improvement. We also find that detecting the intensity of sarcasm is very difficult for neural systems.

Future work could involve adding more data and/or more languages. On the model side, learning joint embedding or early fusion of modalities could be interesting directions. Memotion analysis is a relatively new problem and is far from completion. We hope our work attracts more research attention towards the analysis of memes.

References

[1] R. Dawkins, N. Davis, The selfish gene, Macat Library, 2017.
[2] R. E. Guadagno, D. M. Rempala, S. Murphy, B. M. Okdie, What makes a video go viral? an analysis of emotional contagion and internet memes, Computers in Human

Behavior 29 (2013) 2312–2319. URL: https://www.sciencedirect.com/science/article/pii/S0747563213001192. doi:https://doi.org/10.1016/j.chb.2013.04.016.

[3] M. Moody-Ramirez, A. B. Church, Analysis of facebook meme groups used during the 2016 us presidential election, Social Media + Society 5 (2019) 2056305118808799. URL: https://doi.org/10.1177/2056305118808799. doi:10.1177/2056305118808799. arXiv:https://doi.org/10.1177/2056305118808799.

[4] C. A. Smith, Weaponized iconoclasm in internet memes featuring the expression 'fake news', Discourse & Communication 13 (2019) 303–319. URL: https://doi.org/10.1177/1750481319835639. doi:10.1177/1750481319835639. arXiv:https://doi.org/10.1177/1750481319835639.

[5] R. Al-Natour, The digital racist fellowship behind the anti-aboriginal internet memes, Journal of Sociology 57 (2021) 780–805. URL: https://doi.org/10.1177/1440783320964536. doi:10.1177/1440783320964536. arXiv:https://doi.org/10.1177/1440783320964536.

[6] S. Pramanick, S. Sharma, D. Dimitrov, M. S. Akhtar, P. Nakov, T. Chakraborty, Momenta: A multimodal framework for detecting harmful memes and their targets, arXiv preprint arXiv:2109.05184 (2021).

[7] S. Pramanick, D. Dimitrov, R. Mukherjee, S. Sharma, M. Akhtar, P. Nakov, T. Chakraborty, et al., Detecting harmful memes and their targets, arXiv preprint arXiv:2110.00413 (2021).

[8] L. Yue, W. Chen, X. Li, W. Zuo, M. Yin, A survey of sentiment analysis in social media 60 (2019) 617–663. URL: https://doi.org/10.1007/s10115-018-1236-4. doi:10.1007/s10115-018-1236-4.

[9] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, C. Potts, Learning word vectors for sentiment analysis, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Portland, Oregon, USA, 2011, pp. 142–150. URL: http://www.aclweb.org/anthology/P11-1015.

[10] L.-P. Morency, R. Mihalcea, P. Doshi, Towards multimodal sentiment analysis: Harvesting opinions from the web, ICMI '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 169–176. URL: https://doi.org/10.1145/2070481.2070509. doi:10.1145/2070481.2070509.

[11] A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, in: LREC, 2010.

[12] Z. Kozareva, B. Navarro, S. Vázquez, A. Montoyo, UA-ZBSA: A headline emotion classification through web information, in: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 334–337. URL: https://aclanthology.org/S07-1072.

[13] N. C. Dang, M. N. Moreno-García, F. De la Prieta, Sentiment analysis based on deep learning: A comparative study, Electronics 9 (2020) 483. URL: http://dx.doi.org/10.3390/electronics9030483. doi:10.3390/electronics9030483.

[14] A. Mostafazadeh Davani, D. Kiela, M. Lambert, B. Vidgen, V. Prabhakaran, Z. Waseem (Eds.), Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021), Association for Computational Linguistics, Online, 2021. URL: https://aclanthology.org/2021.woah-1.0.

[15] 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.

[16] 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.

[17] J. Struß, M. Siegel, J. Ruppenhofer, M. Wiegand, M. Klenner, Overview of germeval task 2, 2019 shared task on the identification of offensive language, 2019.

[18] V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. Rangel Pardo, P. Rosso, M. Sanguinetti, SemEval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 54–63. URL: https://aclanthology.org/S19-2007. doi:10.18653/v1/S19-2007.

[19] R. Gomez, J. Gibert, L. Gomez, D. Karatzas, Exploring hate speech detection in multimodal publications, 2019. arXiv:1910.03814.

[20] R. Kumar, A. K. Ojha, S. Malmasi, M. Zampieri, Benchmarking aggression identification in social media, in: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 1–11. URL: https://aclanthology.org/W18-4401.

[21] S. MacAvaney, H.-R. Yao, E. Yang, K. Russell, N. Goharian, O. Frieder, Hate speech detection: Challenges and solutions, PLOS ONE 14 (2019) 1–16. URL: https://doi.org/10.1371/journal.pone.0221152. doi:10.1371/journal.pone.0221152.

[22] M. S. Jahan, M. Oussalah, A systematic review of hate speech automatic detection using natural language processing, 2021. arXiv:2106.00742.

[23] C. Sharma, D. Bhageria, W. Scott, S. PYKL, A. Das, T. Chakraborty, V. Pulabaigari, B. Gamback, Semeval-2020 task 8: Memotion analysis – the visuo-lingual metaphor!, 2020. arXiv:2008.03781.

[24] D. Kiela, H. Firooz, A. Mohan, V. Goswami, A. Singh, P. Ringshia, D. Testuggine, The hateful memes challenge: Detecting hate speech in multimodal memes, 2021. arXiv:2005.04790.

[25] S. Suryawanshi, B. R. Chakravarthi, M. Arcan, P. Buitelaar, Multimodal meme dataset (MultiOFF) for identifying offensive content in image and text, in: Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying, European Language Resources Association (ELRA), Marseille, France, 2020, pp. 32–41. URL: https://aclanthology.org/2020.trac-1.6.

[26] M. Miliani, G. Giorgi, I. Rama, G. Anselmi, G. Lebani, DANKMEMES @ EVALITA 2020: The Memeing of Life: Memes, Multimodality and Politics, 2020, pp. 275–283. doi:10.4000/books.aaccademia.7330.

[27] S. Suryawanshi, B. R. Chakravarthi, Findings of the shared task on troll meme classification in Tamil, in: Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages, Association for Computational Linguistics, Kyiv, 2021, pp. 126–132. URL: https://aclanthology.org/2021.dravidianlangtech-1.16.

[28] Y. Guo, J. Huang, Y. Dong, M. Xu, Guoym at SemEval-2020 task 8: Ensemble-based

classification of visuo-lingual metaphor in memes, in: Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics, Barcelona (online), 2020, pp. 1120–1125. URL: https://aclanthology.org/2020.semeval-1.148. doi:10.18653/v1/2020.semeval-1.148.

[29] G.-A. Vlad, G.-E. Zaharia, D.-C. Cercel, C.-G. Chiru, S. Trausan-Matu, Upb at semeval-2020 task 8: Joint textual and visual modeling in a multi-task learning architecture for memotion analysis, 2020. arXiv:2009.02779.

[30] R. Zhu, Enhance multimodal transformer with external label and in-domain pretrain: Hateful meme challenge winning solution, 2020. arXiv:2012.08290.

[31] S. Ramamoorthy, N. Gunti, S. Mishra, S. S, A. Reganti, P. Patwa, A. Das, T. Chakraborty, A. Sheth, A. Ekbal, C. Ahuja, Memotion 2: Dataset on sentiment and emotion analysis of memes (2022).

[32] T. T. Nguyen, N. T. Pham, H. N. Ngoc Duy Nguyen, L. H. Nguyen, Y.-G. Kim, HCILab at Memotion 2.0 2022: Analysis of sentiment, emotion and intensity of emotion classes from meme images using single and multi modalities, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[33] M. Tan, Q. Le, Efficientnetv2: Smaller models and faster training, in: International Conference on Machine Learning, PMLR, 2021, pp. 10096–10106.

[34] 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).

[35] S. Pramanick, M. S. Akhtar, T. Chakraborty, Exercise? i thought you said 'extra fries': Leveraging sentence demarcations and multi-hop attention for meme affect analysis, 2021. arXiv:2103.12377.

[36] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q. V. Le, Autoaugment: Learning augmentation policies from data, 2019. URL: https://arxiv.org/pdf/1805.09501.pdf.

[37] G. Andrew, R. Arora, J. Bilmes, K. Livescu, Deep canonical correlation analysis, in: S. Dasgupta, D. McAllester (Eds.), Proceedings of the 30th International Conference on Machine Learning, volume 28 of *Proceedings of Machine Learning Research*, PMLR, Atlanta, Georgia, USA, 2013, pp. 1247–1255. URL: https://proceedings.mlr.press/v28/andrew13.html.

[38] B. Duan, Y. Zhu, BROWALLIA at Memotion 2.0 2022 : Multimodal memotion analysis with modified ogb strategies, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[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] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780. URL: https://doi.org/10.1162/neco.1997.9.8.1735. doi:10.1162/neco.1997.9.8.1735.

[41] W. Wang, D. Tran, M. Feiszli, What makes training multi-modal classification networks hard?, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12695–12705.

[42] Y. Zhuang, Y. Zhang, Yet at Memotion 2.0 2022 : Hate speech detection combining bilstm and fully connected layers, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

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

[44] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543. URL: http://www.aclweb.org/anthology/D14-1162.

[45] K. N. Phan, G.-S. Lee, H.-J. Yang, S.-H. Kim, Little Flower at Memotion 2.0 2022 : Ensemble of multi-modal model using attention mechanism in memotion analysis, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[46] A.-M. Bucur, A. Cosma, I.-B. Iordache, BLUE at Memotion 2.0 2022: You have my image, my text and my transformer, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

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

[48] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: International Conference on Machine Learning, PMLR, 2021, pp. 8748–8763.

[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 and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992. URL: https://aclanthology.org/D19-1410. doi:10.18653/v1/D19-1410.

[50] W. Cao, V. Mirjalili, S. Raschka, Rank consistent ordinal regression for neural networks with application to age estimation, Pattern Recognition Letters 140 (2020) 325–331. URL: https://www.sciencedirect.com/science/article/pii/S016786552030413X. doi:https://doi.org/10.1016/j.patrec.2020.11.008.

[51] G. G. Lee, M. Shen, Amazon PARS at Memotion 2.0 2022: Multi-modal multi-task learning for Memotion 2.0 challenge, in: Proceedings of De-Factify: Workshop on Multimodal Fact Checking and Hate Speech Detection, CEUR, 2022.

[52] L. H. Li, M. Yatskar, D. Yin, C.-J. Hsieh, K.-W. Chang, Visualbert: A simple and performant baseline for vision and language, 2019. arXiv:1908.03557.

[53] 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.

[54] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 28, Curran Associates, Inc., 2015. URL: https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.