# Memotion 2: Dataset on Sentiment and Emotion Analysis of Memes

**11 authors**, including:

Nethra Gunti
Indian Institute of Information Technology Sri City

**4** PUBLICATIONS   **0** CITATIONS

Suryavardan S
Indian Institute of Information Technology, Sri City, Chittoor

**4** PUBLICATIONS   **0** CITATIONS

Parth Patwa
Indian Institute of Information Technology Sri City

**33** PUBLICATIONS   **207** CITATIONS

Amitava Das
University of North Texas

**93** PUBLICATIONS   **1,263** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Social Network Communities View project

Project   Emotion Analysis on Suicide Notes View project

# Memotion 2: Dataset on Sentiment and Emotion Analysis of Memes

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

[1]IIIT Sri City, India
[2]Amazon, USA
[3]University of California Los Angeles, 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 commonly used in social media platforms for humour. Generally, memes consist of an image and embedded text. Memes can be used to spread hate or misinformation, hence it is important to study them. The Memotion task [1] conducted at SemEval 2020, released a data of 10k memes annotated with sentiment label (task A), emotion label (task B) and emotion intensity label (task C). It received $\approx$ 30 submissions and 27 papers. However, the best f1 scores were only 0.35, 0.51 and 0.32 respectively, which shows the need for more extensive research along this direction. In this paper, we release a new dataset, Memotion 2 which has 10k annotated memes along the same directions as Memotion 1. We also provide baseline systems for our data.

## 1. Introduction

Memes, usually created with image and/or text, have become a highly popular medium of communication on the internet. Over the past decade, memes have become an integral part of the internet culture, giving communities the power to make their voices heard to large audiences in a matter of few hours. Hence, to understand a community's opinions and alignment with their causes, a good start would is to be conscious of the memes shared by the community [2] .

Fig 1 shows the rapid growth in using memes over the past few years and its success can be attributed to the increased usage of internet by people from all walks of life. Many memes are used for just fun and games, but they are also a powerful medium to voice a community's opinion (or alignment with a cause). With the freedom to express opinion anonymously comes

---

*Equal contribution.

**Figure 1:** The rise of in popularity of internet memes in recent years. (Source: Google trends [3].)

a varying definition of "normal". As difficult as it already is to classify content from unimodal data like simple texts/tweets; including multi-modal data (eg. memes) into the mix only further complicates the problem, given the fact that they are highly unconventional and don't have a fixed template or modality.

A flip side of this powerful medium is that it can be misused to spread hatred in the community. The growing trends in hate speech on social media commensurate with the increased quotidian usage of memes. Considering the effects hate speech might have on individuals, detecting and preventing such content from being spread is important. Williams et al. 2016 [4] is one such work that studies the effect of racist memes on real life and vice versa. It only strengthens the need to thoroughly analyse memes. Detecting hate speech in online platforms has gained traction over the past few years in the research community. Building a system that can work on memes is highly complex because of the nature of multi-modal data in them, the worst-case scenario being that the modalities involved can be complementary and still convey meaning.

Humans show various emotions like rage, disgust, grief, serenity, fear, etc. and each exhibits them in varying levels of intensity. Our dataset, **Memotion 2.0**, focuses on quantifying emotion and their intensities into discrete labels. It also has labels corresponding to the sentiment of a meme. Memotion 2.0 adds to the previous iteration (Memotion 1.0) by providing another set of 10k memes from various social media websites. In this iteration, the collected memes are more widespread in terms of topics ranging from history to world wars, politics, etc. Popular social media sites like Reddit, Facebook, Imgur, Instagram were used as sources for memes. These websites consist of individuals from different countries, religions and ethnic groups. Hence, we

believe that this enables the analysis of multi-modal memes with respect to sentiment analysis and emotion detection using our dataset more closer to general human perception.

The paper is organised as follows: We describe the related work and the task in section 2 and 3 respectively; Section 4 contains the details of the dataset we collected for memotion analysis: Memotion 2.0 ; followed by a brief description of baseline models 5 and their results in 6. We conclude with the mention of future work and limitations, in section 7.

## 2. Related Work

Past years have seen a lot of work related to analysing social media content and detecting emotions, profanity and other such attributes. Majority of hate speech datasets are of textual modality [5, 6] and several of them are from twitter [7, 8, 9].

Due to their reliance on identifying n-grams, phrases or textual patterns, these datasets do not account for subjective bias [10] and may present a lack of context [11, 12] when classifying uni-modal data. The same goes for text based sentiment analysis and emotion analysis datasets such as [13, 14, 15].

The growing ubiquity of Internet memes on social media platforms suggests that it is more than important to consider such multi-modal content. MMHS150K [16] is a dataset collected from Twitter using Hatebase terms. It contains 150,000 tweets and images manually annotated into six classes based on the type of hate speech. Haoti et. al. [17] and Hosseinmardi et. al. [18] provide annotated datasets from Instagram with posts and comments targeted towards combating cyberbullying. Sentiment analysis datasets such as [19] and [20] classify videos or image-text pairs into "positive" or "negative" labels.

While there are several datasets that facilitate computation of social media data, analysing memes has received relatively less attention. MultiOFF [21] is an annotated dataset with 743 memes from Kaggle. While it does have image and text captions, the dataset only has binary labels i.e. "offensive" and "non-offensive". Another notable dataset is hateful memes dataset by facebook [22]. The memes are from social media groups in the United States and they were annotated using a specific definition of hate speech. Each meme can have multiple labels and the labels are defined for multimodal and unimodal hate speech separately. The dataset is of size 10k but has some reconstructed i.e. artificial memes. The Memotion 1.0 task at SemEval 2020 [1] succeeded at drawing attention to the analysis and detection of sentiment and hate speech in memes. The participants were provided with around 10K memes with multiple human annotated labels for 3 tasks - sentiment analysis, emotion analysis, emotion intensity classification. The task achieved a highest F1 score of 0.35, 0.51 and 0.32 for the three tasks respectively.

## 3. Memotion 2.0 task

We release a dataset of 10k annotated memes. Each data point consists of of an image and text associated with it along with labels for each sub-tasks. Similar to Memotion 1 [1], We consider sentiment, emotions and their intensities and hence they form our sub tasks, which are as follows:

**Figure 2:** Annotator Interface. The annotators see a meme and have to mark the sentiment and emotion intensities of the meme. They also have to tell on what basis was the annotation done (text/image/both).



**Figure 3:** Example for Task A. People found this meme to have a negative sentiment. The meme does seem to convey that the person who made this meme wants to stand on a place where the Chinese rocket will crash. Hence, they might have labeled it as having a negative sentiment.

- **Task A: Sentiment Analysis** - Given an Internet meme, the first task is to classify it as a positive, negative or neutral meme. This helps one to understand the sentiment of a meme. Figure 3 explains why a particular meme might have a negative sentiment.
- **Task B: Emotion Classification** - Given an Internet meme, the system has to identify

**Figure 4:** Example for Task B and C. Majority of annotators found this meme's humour intensity as hilarious, sarcasm level as little sarcastic ( maybe because it shows how ineffective online classes are ), not offensive and not motivational. The corresponding labels for Task B will be funny, sarcastic, not offensive and not motivational.

the type of emotion expressed. The categories 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** - The third task is to quantify the extent to which a particular emotion is being expressed. Fig 2 mentions about intensities of each emotion.

  Tasks B and C can be clearly understood by looking at the meme in Figure 4.

## 4. Dataset

In this section we describe the data collections and data annotation process along with a brief summary of the data distribution.

### 4.1. Data Collection

As established in prior sections, memes are highly complex form of data, and it is necessary that we collect them from a wide range of categories. We shortlisted several topics of interests: like politics, religion, sports etc.- and manually downloaded the memes. We've also used a Selenium based web-crawler for a part of data collection, followed by extensive cleaning of the data. All the memes have been collected from public domains, and in order to avoid copyright claims, the source-urls have been attributed in the dataset. We used the Google Vision API[1] to extract the OCR text from the images.

---

[1]https://cloud.google.com/vision
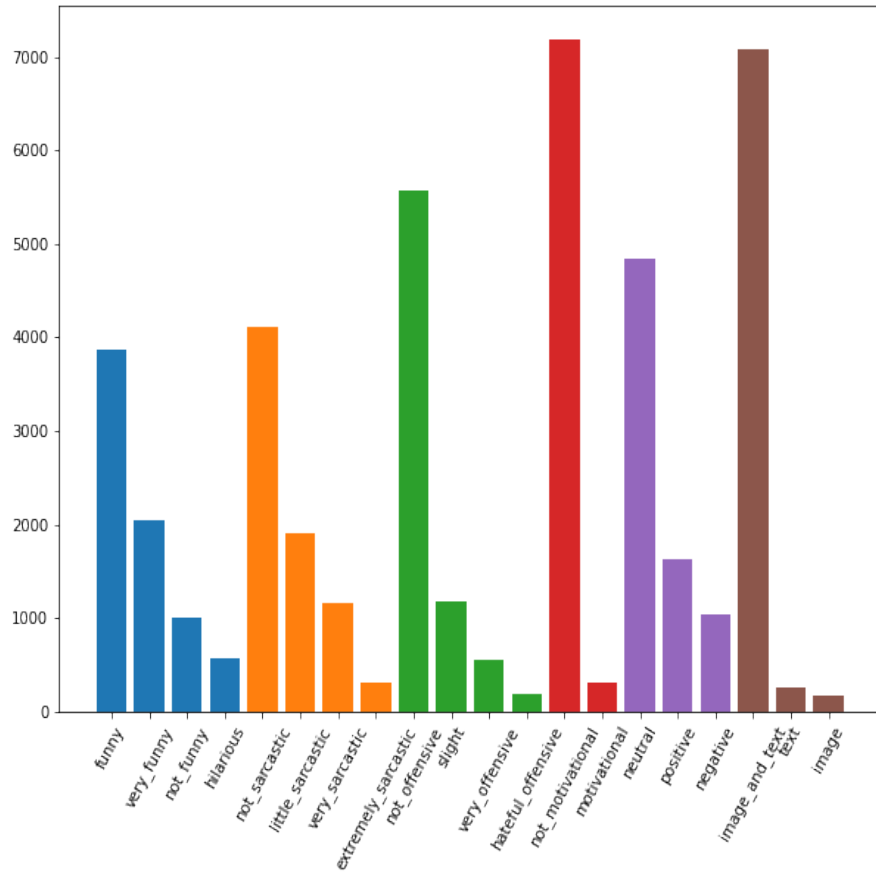
**Figure 5:** Distribution of the samples across all labels. Blue corresponds to 'humour': ['not_funny', 'funny', 'very_funny', 'hilarious']; Orange corresponds to 'sarcasm':['not_sarcastic', 'little_sarcastic', 'very_sarcastic', 'extremely_sarcastic']; Green corresponds to 'offense':['not_offensive', 'slight_offensive', 'very_offensive', 'hateful_offensive']; Red corresponds to 'motivation':['not_motivational', 'motivational']; Violet corresponds to 'overall_sentiment':['negative', 'neutral', 'positive']; Brown corresponds to 'classification_based_on':['image', 'text', 'image_and_text'];

## 4.2. Data Annotation

After collected the data, we turned to Amazon Mechanical Turk(AMT) [2] workers to get it annotated. An annotator has to choose from the following and annotate each meme for all the said fields. For this purpose, they use an interface built by us, as shown in Fig 2. For task A, the annotators were asked to judge what the person who created the meme intended it to be (positive/negative/neutral). For task B and C, the annotators were asked to mark their opinions on the emotion of the meme. The true affect of a meme on an individual depends on their perception of several aspects within the society, and could vary a lot from another individual. We solve this problem with each meme being annotated by 3 different workers. Based on the

---

**Overall Dataset % Distribution**

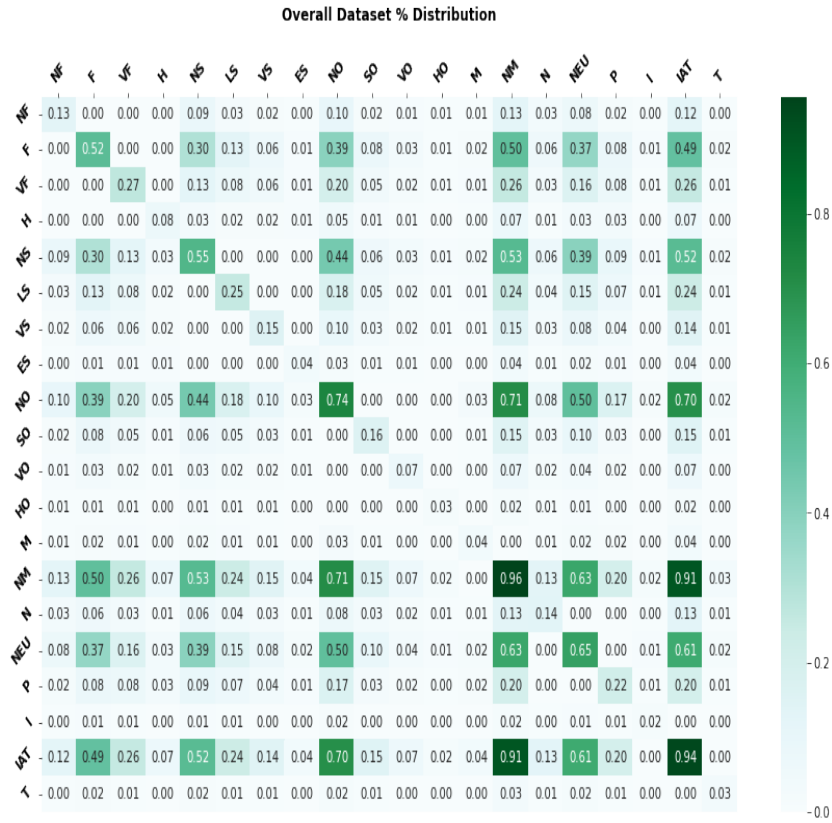| | NF | F | VF | H | NS | LS | VS | ES | NO | SO | VO | HO | M | NM | N | NEU | P | I | IAT | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NF | 0.13 | 0.00 | 0.00 | 0.00 | 0.09 | 0.03 | 0.02 | 0.00 | 0.10 | 0.02 | 0.01 | 0.01 | 0.01 | 0.13 | 0.03 | 0.08 | 0.02 | 0.00 | 0.12 | 0.00 |
| F | 0.00 | 0.52 | 0.00 | 0.00 | 0.30 | 0.13 | 0.06 | 0.01 | 0.39 | 0.08 | 0.03 | 0.01 | 0.02 | 0.50 | 0.06 | 0.37 | 0.08 | 0.01 | 0.49 | 0.02 |
| VF | 0.00 | 0.00 | 0.27 | 0.00 | 0.13 | 0.08 | 0.06 | 0.01 | 0.20 | 0.05 | 0.02 | 0.01 | 0.01 | 0.26 | 0.03 | 0.16 | 0.08 | 0.01 | 0.26 | 0.01 |
| H | 0.00 | 0.00 | 0.00 | 0.08 | 0.03 | 0.02 | 0.02 | 0.01 | 0.05 | 0.01 | 0.01 | 0.00 | 0.00 | 0.07 | 0.01 | 0.03 | 0.03 | 0.00 | 0.07 | 0.00 |
| NS | 0.09 | 0.30 | 0.13 | 0.03 | 0.55 | 0.00 | 0.00 | 0.00 | 0.44 | 0.06 | 0.03 | 0.01 | 0.02 | 0.53 | 0.06 | 0.39 | 0.09 | 0.01 | 0.52 | 0.02 |
| LS | 0.03 | 0.13 | 0.08 | 0.02 | 0.00 | 0.25 | 0.00 | 0.00 | 0.18 | 0.05 | 0.02 | 0.01 | 0.01 | 0.24 | 0.04 | 0.15 | 0.07 | 0.01 | 0.24 | 0.01 |
| VS | 0.02 | 0.06 | 0.06 | 0.02 | 0.00 | 0.00 | 0.15 | 0.00 | 0.10 | 0.03 | 0.02 | 0.01 | 0.01 | 0.15 | 0.03 | 0.08 | 0.04 | 0.00 | 0.14 | 0.01 |
| ES | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.04 | 0.03 | 0.01 | 0.01 | 0.00 | 0.00 | 0.04 | 0.01 | 0.02 | 0.01 | 0.00 | 0.04 | 0.00 |
| NO | 0.10 | 0.39 | 0.20 | 0.05 | 0.44 | 0.18 | 0.10 | 0.03 | 0.74 | 0.00 | 0.00 | 0.00 | 0.03 | 0.71 | 0.08 | 0.50 | 0.17 | 0.02 | 0.70 | 0.02 |
| SO | 0.02 | 0.08 | 0.05 | 0.01 | 0.06 | 0.05 | 0.03 | 0.01 | 0.00 | 0.16 | 0.00 | 0.00 | 0.01 | 0.15 | 0.03 | 0.10 | 0.03 | 0.00 | 0.15 | 0.01 |
| VO | 0.01 | 0.03 | 0.02 | 0.01 | 0.03 | 0.02 | 0.02 | 0.01 | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 | 0.07 | 0.02 | 0.04 | 0.02 | 0.00 | 0.07 | 0.00 |
| HO | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.02 | 0.01 | 0.01 | 0.00 | 0.00 | 0.02 | 0.00 |
| M | 0.01 | 0.02 | 0.01 | 0.00 | 0.02 | 0.01 | 0.01 | 0.00 | 0.03 | 0.01 | 0.00 | 0.00 | 0.04 | 0.00 | 0.01 | 0.02 | 0.02 | 0.00 | 0.04 | 0.00 |
| NM | 0.13 | 0.50 | 0.26 | 0.07 | 0.53 | 0.24 | 0.15 | 0.04 | 0.71 | 0.15 | 0.07 | 0.02 | 0.00 | 0.96 | 0.13 | 0.63 | 0.20 | 0.02 | 0.91 | 0.03 |
| N | 0.03 | 0.06 | 0.03 | 0.01 | 0.06 | 0.04 | 0.03 | 0.01 | 0.08 | 0.03 | 0.02 | 0.01 | 0.01 | 0.13 | 0.14 | 0.00 | 0.00 | 0.00 | 0.13 | 0.01 |
| NEU | 0.08 | 0.37 | 0.16 | 0.03 | 0.39 | 0.15 | 0.08 | 0.02 | 0.50 | 0.10 | 0.04 | 0.01 | 0.02 | 0.63 | 0.00 | 0.65 | 0.00 | 0.01 | 0.61 | 0.02 |
| P | 0.02 | 0.08 | 0.08 | 0.03 | 0.09 | 0.07 | 0.04 | 0.01 | 0.17 | 0.03 | 0.02 | 0.00 | 0.02 | 0.20 | 0.00 | 0.00 | 0.22 | 0.01 | 0.20 | 0.01 |
| I | 0.00 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.01 | 0.01 | 0.02 | 0.00 | 0.00 |
| IAT | 0.12 | 0.49 | 0.26 | 0.07 | 0.52 | 0.24 | 0.14 | 0.04 | 0.70 | 0.15 | 0.07 | 0.02 | 0.04 | 0.91 | 0.13 | 0.61 | 0.20 | 0.00 | 0.94 | 0.00 |
| T | 0.00 | 0.02 | 0.01 | 0.00 | 0.02 | 0.01 | 0.01 | 0.00 | 0.02 | 0.01 | 0.00 | 0.00 | 0.00 | 0.03 | 0.01 | 0.02 | 0.01 | 0.00 | 0.00 | 0.03 |

**Figure 6:** Overall Distribution of the dataset showing overlap between all 20 labels. [NF: not_funny; F: funny; VF: very_funny; H: hilarious; NS: not_sarcastic; LS: little_sarcastic; VS: very_sarcastic; ES: extremely_sarcastic; NO: not_offensive; SO: slight_offensive; VO: very_offensive; HO: hateful_offensive; M: motivational; NM: not_motivational; N: negative; NEU: neutral; P: positive; I: image; IAT: image_and_text; T: text].

majority voting scheme, the final annotations are adjudicated.

## 4.3. Data Distribution

The dataset consists of 10,000 images divided into a train-val-test split with 8500-1500-1500 images. Each meme is annotated for its Overall Sentiment (positive, neutral, negative), Emotion (humour, sarcasm, offense, motivation) and Scale of Emotion (0-4 levels). Along with the said attributes, we also introduce a new attribute called "classification_based_on", which denotes if the annotation has been made on the basis of image only, text only, or both image and text data. Fig.5. shows the distribution of memes across all the 20 labels.

The statistical summaries in Fig. 6 and Fig.7 show the overlapping emotions in memes, which proves the aforementioned challenges. Several interesting points can be inferred from the tables like many offensive memes are funny. It can also be observed that many of the memes are funny and non-motivational.
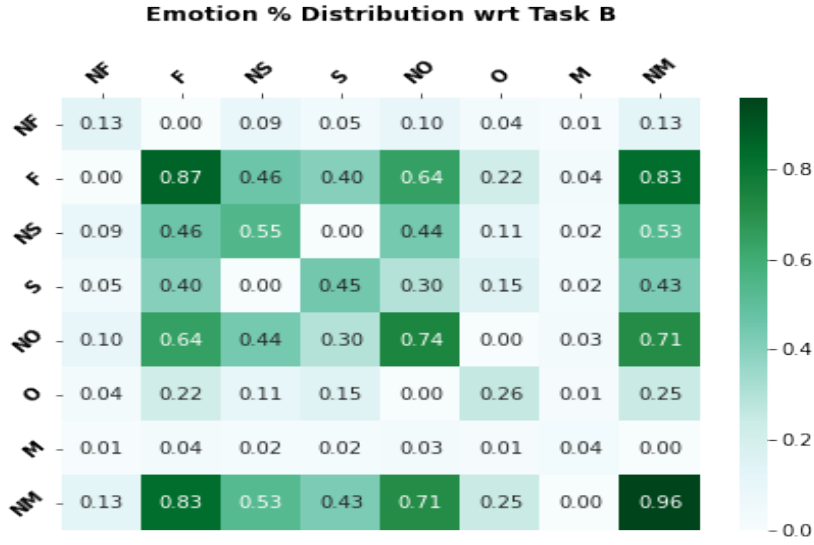
**Figure 7:** Distribution of the dataset showing overlap between emotion labels for Task B. [NF: not_funny; F: funny; NS: not_sarcastic; S: sarcastic; NO: not_offensive; O: offensive; M: motivational; NM: not_motivational].

## 5. Baseline Model

In this section, we describe the various baseline models that we tried.

### 5.1. Text Model

Memes are known to convey information with both image and text. However, it is noticed that many memes have the same template (image) but different text and by implication, different messages to convey. Recognizing of the emotion induced in such memes would require accurate modelling of the textual influence. For this purpose, we evaluate the affect of textual features using LSTM. The additional controlling knobs in LSTM, makes it more suitable than simple RNNs for these classification tasks. Table 1. shows the baseline Weighted F1 scores of the model on all three memotion tasks.

### 5.2. Vision + Text Model

Experiments are carried out considering the multi-modal features of memes i.e both image and OCR text. BERT is a widely known attention model that provides State-Of-The-Art results on various text related tasks. We make use of BERT to extract features from the OCR text. The text and image features are represented by the CLS output of BERT and the final MLP layer output of ResNet-50 respectively. We concatenate the obtained features and use two layers of MLP for classification. Scores on Tasks A, B and C are reported in table 1.

## 6. Results

Baseline results in Table 1 show Weighted F1 scores for each task and sub-task. It can inferred from the table that multi-modal Image+Text models with scores 43.90%, 73.72% and 51.08%, perform better than the Text-Only models with scores 28.29%, 31.38%, 50.94% for Task A, Task B, and Task C, respectively. The table shows the inconsistency of performance throughout Task B and Task C sub-tasks upon using Text-Only model, unlike Image+Text model. Although it can't be concluded, one can clearly see the importance of using multi-modal data for the said tasks. This dataset will be publicly available and we leave it to the future works, to come up with novel methods which dig deeper into Memotion Analysis and provide statistical insights to multi-modal relations of a Meme.

| Task | Class | Weighted F1 score | |
|---|---|---|---|
| | | Text-Only | Image+Text |
| Task-A | Sentiment | 28.29% | 43.39% |
| | Humour | 52.07% | 78.78% |
| | Sarcasm | 37.80% | 64.43% |
| Task-B | Offensive | 24.57% | 55.17% |
| | Motivation | 11.09% | 95.95% |
| | Average | 31.38% | 73.58% |
| | Humour | 45.73% | 33.23% |
| | Sarcasm | 43.67% | 25.88% |
| Task-C | Offensive | 48.37% | 49.14% |
| | Motivation | 65.97% | 95.95% |
| | Average | 50.94% | 51.05% |

**Table 1**
Baseline scores (Weighted F1) of Text-Only models and multi-modal Image+Text models on Memotion Analysis tasks. Results show that Image+Text models perform better than Text-Only models, including the latter being inconsistent over all.

## 7. Conclusion

In this paper, we introduce a novel task of identifying memes and classifying them using a multimodal setting. To the best of our knowledge, this is the first large-scale multimodal dataset for meme classification. In order to provide a fine-grained and extensive analysis of tweets We provide gold-data for three different verticals namely- sentiment analysis, emotion classification and intensity of emotion. We also provide text only and multimodal baselines for each of these tasks. While the text-only model uses LSTM, the multi-modal model uses ResNet-50 + BERT, which is a recent SOTA model on many popular image-text tasks like captioning, VQA, phrase grounding etc. The performance of these models indicate that incorporating both images and text for all the tasks improves performance, however, it must be noted that our models are only preliminary and more innovative methods will improve performance furthermore. In the future, we intend to extend our work by releasing datasets for memes of other languages, identification of targets of hate and diffusion patterns etc.

# References

[1] C. Sharma, D. Bhageria, W. Paka, Scott, S. P Y K L, A. Das, T. Chakraborty, V. Pulabaigari, B. Gambäck, SemEval-2020 Task 8: Memotion Analysis-The Visuo-Lingual Metaphor!, in: Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020), Association for Computational Linguistics, Barcelona, Spain, 2020.

[2] N. Gal, L. Shifman, Z. Kampf, "it gets better": Internet memes and the construction of collective identity, New Media & Society 18 (2016) 1698–1714. URL: https://doi.org/10.1177/1461444814568784. doi:10.1177/1461444814568784. arXiv:https://doi.org/10.1177/1461444814568784.

[3] Google, Google search interest in "memes", 2022. URL: https://trends.google.com/trends/explore?date=all&q=memes.

[4] A. Williams, C. Oliver, K. Aumer, C. Meyers, Racial microaggressions and perceptions of internet memes, Computers in Human Behavior 63 (2016) 424–432. doi:10.1016/j.chb.2016.05.067.

[5] 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.

[6] J. Qian, A. Bethke, Y. Liu, E. Belding, W. Y. Wang, A benchmark dataset for learning to intervene in online hate speech, 2019. arXiv:1909.04251.

[7] Z. Waseem, D. Hovy, Hateful symbols or hateful people? predictive features for hate speech detection on twitter, 2016, pp. 88–93. doi:10.18653/v1/N16-2013.

[8] Z. Waseem, Are you a racist or am i seeing things? annotator influence on hate speech detection on twitter, in: NLP+CSS@EMNLP, 2016.

[9] P. Burnap, M. Williams, Us and them: identifying cyber hate on twitter across multiple protected characteristics, EPJ Data Science 5 (2016). doi:10.1140/epjds/s13688-016-0072-6.

[10] T. Davidson, D. Warmsley, M. Macy, I. Weber, Automated hate speech detection and the problem of offensive language, 2017. arXiv:1703.04009.

[11] T. Davidson, D. Bhattacharya, I. Weber, Racial bias in hate speech and abusive language detection datasets, 2019. arXiv:1905.12516.

[12] M. Sap, D. Card, S. Gabriel, C. Yejin, N. Smith, The risk of racial bias in hate speech detection, 2019, pp. 1668–1678. doi:10.18653/v1/P19-1163.

[13] 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.

[14] A. Go, R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, Processing 150 (2009).

[15] P. Patwa, G. Aguilar, S. Kar, S. Pandey, S. PYKL, B. Gambäck, T. Chakraborty, T. Solorio, A. Das, SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets, in: Proceedings of the Fourteenth Workshop on Semantic Evaluation, International Committee for Computational Linguistics, Barcelona (online), 2020. URL: https://aclanthology.org/2020.semeval-1.100.

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

[17] H. Zhong, H. Li, A. Squicciarini, S. Rajtmajer, C. Griffin, D. Miller, C. Caragea, Content-driven detection of cyberbullying on the instagram social network, 2016.

[18] H. Hosseinmardi, S. A. Mattson, R. I. Rafiq, R. Han, Q. Lv, S. Mishra, Detection of cyberbullying incidents on the instagram social network, 2015. arXiv:1503.03909.

[19] L.-P. Morency, R. Mihalcea, P. Doshi, Towards Multimodal Sentiment Analysis: Harvesting Opinions from The Web, in: International Conference on Multimodal Interfaces (ICMI 2011), Alicante, Spain, 2011. URL: http://ict.usc.edu/pubs/Towards%20Multimodal%20Sentiment%20Analysis-%20Harvesting%20Opinions%20from%20The%20Web.pdf.

[20] A. Hu, S. Flaxman, Multimodal sentiment analysis to explore the structure of emotions, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery |& Data Mining (2018). URL: http://dx.doi.org/10.1145/3219819.3219853. doi:10.1145/3219819.3219853.

[21] 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.

[22] 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.