

Exercise? I thought you said 'Extra Fries': Leveraging Sentence Demarcations and Multi-hop Attention for Meme Affect Analysis

Shraman Pramanick, Md. Shad Akhtar, Tanmoy Chakraborty
IIIT-Delhi

ICWSM-2021
June 7-10, 2021

Motivation

- Interpreting memes is a challenging task often because of the implicit world knowledge and overlapping information.
- Two memes can have same image (and vice versa) but can convey different semantics based on the other modality.
- World knowledge plays a crucial part in establishing a relationship between a meme and its reference situation.

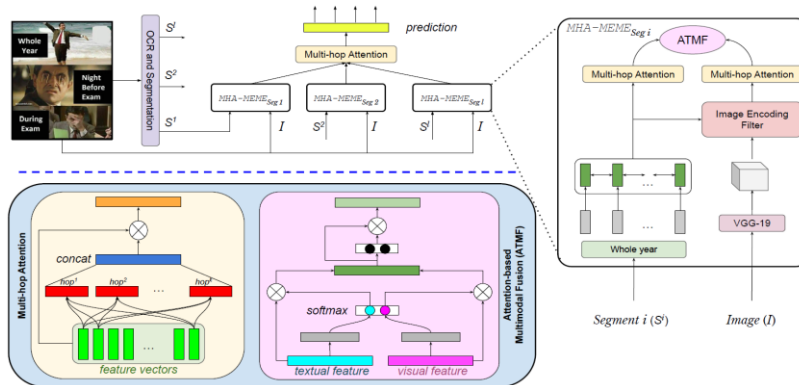


- Annotating memes is challenging, as emotion about memes highly depends upon individual's perception of an array of aspects within society - **"Subjective Perception Problem"** (Zhao et al. 2018)

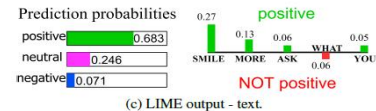
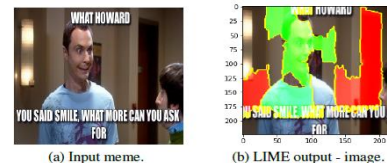
Objective and Contributions

- We propose multi-hop attention-based framework, **MHA-Meme**, which captures complementary features from meme image and text.
- We evaluate MHA-Meme on the **'Memotion Analysis'** dataset for all three sub-tasks -sentiment classification, affect classification, and affect class quantification and report benchmark results.
- Memotion Analysis' dataset contains 6601 training samples and 1879 test samples. Additionally, we to validate the generalizability of MHA-Meme, we collected and annotated an additional set of 334 memes.
- Sentiment Classes** – [positive, neutral, negative]
- Affect Classes** – [humor, sarcasm, offense, motivation]
- We establish the **interpretability of MHA-Meme** using LIME framework

MHA-Meme Architecture



Explainability of MHA-Meme



Segment: WHAT HOWARD
 Segment: YOU SAID SMILE WHAT MORE CAN YOU ASK FOR

(d) Attention weights as computed by MHA-Meme.

Results

| Models | Sentiment classification | | | | Affect classification (Avg) | | | | Affect quantification (Avg) | | | | |
|--------|-----------------------------------|-------------------|-------------------|-------------------|-----------------------------|-------------------|-------------------|-------------------|-----------------------------|-------------------|-------------------|-------------------|-------|
| | Macro F1 | | Micro F1 | | Macro F1 | | Micro F1 | | Macro F1 | | Micro F1 | | |
| | Test _A | Test _B | Test _A | Test _B | Test _A | Test _B | Test _A | Test _B | Test _A | Test _B | Test _A | Test _B | |
| T | BiLSTM - OCR | 0.338 | 0.373 | 0.509 | 0.572 | 0.421 | 0.455 | 0.542 | 0.570 | 0.302 | 0.310 | 0.420 | 0.438 |
| | BERT - OCR | 0.336 | 0.375 | 0.512 | 0.570 | 0.422 | 0.449 | 0.549 | 0.571 | 0.295 | 0.298 | 0.395 | 0.402 |
| | BiLSTM - OCR _{Seg} | 0.352 | 0.391 | 0.560 | 0.594 | 0.475 | 0.490 | 0.570 | 0.594 | 0.319 | 0.332 | 0.422 | 0.442 |
| | BERT - OCR _{Seg} | 0.351 | 0.384 | 0.538 | 0.580 | 0.471 | 0.482 | 0.563 | 0.581 | 0.311 | 0.316 | 0.418 | 0.425 |
| I | InceptionV3 | 0.322 | 0.358 | 0.516 | 0.557 | 0.407 | 0.430 | 0.499 | 0.525 | 0.288 | 0.287 | 0.402 | 0.406 |
| | V16 | 0.318 | 0.355 | 0.521 | 0.560 | 0.399 | 0.432 | 0.505 | 0.532 | 0.286 | 0.295 | 0.411 | 0.418 |
| | V19 | 0.325 | 0.367 | 0.525 | 0.562 | 0.413 | 0.448 | 0.518 | 0.550 | 0.292 | 0.300 | 0.405 | 0.419 |
| T+I | BERT - OCR _{Seg} + V19 | 0.356 | 0.410 | 0.585 | 0.624 | 0.508 | 0.529 | 0.620 | 0.645 | 0.325 | 0.362 | 0.424 | 0.435 |
| | BiLSTM - OCR _{Seg} + V19 | 0.376 | 0.426 | 0.608 | 0.635 | 0.523 | 0.545 | 0.682 | 0.698 | 0.333 | 0.360 | 0.430 | 0.444 |

Comparative study against baseline & various SOTAs



References:

- Zhao, S.; Ding, G.; Huang, Q.; Chua, T.-S.; Schuller, B. W.; and Keutzer, K. 2018. Affective Image Content Analysis: A Comprehensive Survey. In IJCAI, 5534–5541.
- Sharma, C.; 2020. SemEval-2020 Task 8: Memotion Analysis- the Visuo-Lingual Metaphor! In SemEval-2020, 759–773.
- Lee, K.-H.; Chen, X.; Hua, G.; Hu, H.; and He, X. 2018. Stacked cross attention for image-text matching. In ECCV, 201–216.

