Soumil Gupta
2022102035
soumil.gupta@students.iiit.ac.in

1 Summary

The research paper discusses the difficulty in comprehending the context and background of memes due to the lack of effective methods for dynamically interpreting their context. It introduces a new task named MEMEX, designed to unearth the context explaining a meme's background when provided with the meme and a corresponding document.

Furthermore, the paper outlines how MEMEX could be beneficial for content moderation on social media platforms, especially during critical times such as elections or health crises.

To facilitate this, the authors have created a new dataset known as MCC and developed a multimodal neural architecture called MIME. MIME is designed to assess if a meme and its associated context are logically connected. The architecture of MIME includes several elements, such as a text encoder to process words, a multimodal encoder, an advanced meme encoder enriched with external knowledge, a Transformer model that is aware of meme-specific nuances, and a Long Short-Term Memory (LSTM) model tailored for memes.

For encoding textual content from potential contexts, MIME utilizes a pre-trained BERT model, while for capturing a multimodal understanding of the meme, it employs a pre-trained Multimodal Bitransformers for Multimodal Representations (MMBT) model. Additionally, to enhance the meme's representation with broader knowledge, MIME uses a pre-trained graph convolutional network (GCN) that is trained on ConceptNet, a knowledge graph.

2 Strengths of the paper

Here are some of the strengths of the paper that I found to be very important:

- Memes often hide their true meanings, showing how useful MEMEX can be for a wide range of applications that combine pictures and words but don't clearly align. This includes finding related content for different art types, news photos, and creative images used in digital marketing.
- The paper introduces a new dataset named MCC (Meme Context Corpus) due to the lack of large, publicly available datasets that contain memes and their background information. This dataset specifically tackles the issues arising from the lack of knowledge and the complexity of combining different modes of information, ensuring better identification of related contexts for memes.
- MIME works without needing manually saved details or metadata. It assumes that the related context can be easily found online.
- MIME performs better than many other systems that either use only one mode of information or multiple modes, showing a significant improvement of about 4% in F1-score compared to the next best system.

3 Weaknesses of the paper

It is important to realize that not everything is perfect. The paper has a few critical weaknesses that need to be addressed. Here are three points that highlight this aspect:

- The ability to identify relevant context could enable wrongdoers to subtly embed harmful messages within memes, aiming to deceive content moderators. These sophisticatedly crafted memes might not easily reveal appropriate context, allowing these individuals to bypass moderation efforts.

- MEMEX faces challenges in accurately recognizing and understanding complex memes, shown through instances of no predictions, partial matches, and incorrect guesses. These issues arise from the complexity of memes' meanings, difficulties in combining visual facts with text, not enough textual clues to learn contextual links, and the chance of incorrectly identifying unrelated context due to biases in language.
- The dataset could be misused for targeting specific groups, individuals, or organizations in a biased manner, potentially involving sensitive demographic information. To prevent misuse, human oversight is necessary.

4 Suggestions for Improvements to the paper

There is always a scope for improvement, the same is true for this paper. The following three suggestions, if implemented, could help make the model much more relevant and robust:

- The authors have mainly focused on memes related to politics and history in English. Expanding the model to cover more subjects beyond politics and history, and including more languages, would make it useful to a broader audience.
- The dataset creation process, specifically the annotation process, relied on two paid annotators, which makes scaling the model costly. Allowing the open-source community to contribute to annotating, with proper checks, could speed up the process and reduce financial constraints.
- An area for enhancement is the integration of visual and textual information in memes, which is currently not as effective as it could be in MIME. Improving how the model processes and recognizes visuals and text related to memes could enhance its accuracy and performance.