MEME*X*: Detecting Explanatory Evidence for Memes via Knowledge-Enriched Contextualization

Name: Madadapu HemanthSai
College: MLR institute of technology
Mail: hemanthsai826@gmail.com

1. Summary of the Paper:

The entire paper revolves around the concept of dynamic extraction of visual and linguistic details from the memes. Such details contribute to decode simple messages hidden in them. Indirectly, it helps to understand and gain insights into societal, cultural, and political perspectives across different social media platforms. The entire research can be summed up by a research question stated as follows.

Q-1. How to detect explanatory evidence for memes using knowledge enriched contextualization?

1.1 Hypothesis tested:

The following hypothesis was tested aligning with the research question of the paper.

- If a model uses knowledge-enriched contextualization, it is more effective in identifying explanatory evidence for memes than a model that does not use this approach.

1.2 How the hypothesis was tested:

- The hypothesis was tested by developing and evaluating the performance of MIME framework whose F1 scores were compared with unimodal and multimodal systems working on the same context without using knowledge-enriched contextualization approach.

1.3 Components of the proposed system:

- MEMEX - It is a task used to mine the information from the meme when it is given along with a related document.
- MIME (MultImodal Meme Explainer) - MIME is the heart of the entire system. It is a multimodal neural framework which uses a layered approach and common sense-enriched meme representation to capture the semantic dependencies between the meme and the context. MIME contains several sub-components.
    - Visual Encoder: It is responsible to extract the features from the meme (such as colours, shapes, objects etc.) and encode them.
    - Textual Encoder: It is responsible to encode the text in the meme image, performing semantic operations to identify the dependencies.

    - Knowledge Encoder: It is responsible to integrate the external knowledge resources into meme representation.
    - Integrator: It is responsible to integrate the encoded meme and text into single representations to explain the semantic relationships at different levels of abstraction and produce an explanation as output.
    - Decision layer: It is responsible to classify the generated output as explanatory or non-explanatory for the meme.
- MCC (Meme Context Corpus) - It is the dataset specifically created for the MEMEX task. It contains paired memes and its textual explanations for training the model.

1.4 Findings

- **Kind of study:** Empirical analysis: The study evaluates the effectiveness on real-world data.
- **Type of data used**: Quantitative: Both the MCC dataset and the evaluation metrics used, i.e., F1 score are quantitative.
- **Methodology employed:** experimental and computational (quantitative in nature).
- **Type of ML algorithm used:** Supervised Machine Learning algorithm
- **Attributes under observation:** semantic relationships and their levels of abstraction, multimodality, and knowledge integration
- **Baseline models used**: Unimodal (BERT, ViT), Multimodal (Early-Fusion, MMBT, CLIP, BAN, VisualBERT).

2. Working

Using the MCC dataset, the MEMEX algorithm is aimed to understand the background and complex meanings inside memes. It makes use of the MIME framework, which applies a multimodal neural method to capture the links between a meme and its context. It modulates and combines representations from the meme and contextual information using a Gated Multimodal Fusion block. This system generates key and value vectors based on the meme representation and regulates their interaction using gating parameters. MEMEX beats competing unimodal and multimodal systems in evaluations, highlighting its superior performance in meme analysis.

- **Independent Variable**: (1) Features extracted from the input (meme and its respective context) given to the system. (2) External knowledge databases.
- **Dependent Variable:** Accuracy of MIME in identifying the explanatory sentences for a given meme measured using metrics like F1 score.

3. Results and interpretations

- The suggested MIME system outperforms many unimodal and multimodal systems, achieving a 4% F1-score improvement over the best baseline.
- Average performance = Unimodal systems (BERT being the best and ViT being the worst); Multimodal systems either competed or outperformed the unimodal systems.
- In-depth assessments of MIME's performance are carried out, highlighting elements that may contribute to effective modelling of cross-modal contextual connections. These evaluations provide information for better meme comprehension and context mining.
- The empirical investigation and comprehensive ablation study indicate that MIME and its parts are effective. The paper also outlines MIME's proper contextual mapping heuristics, as well as its limits, and suggests potential areas for development.

4. Future Work

- Improving the modeling of cross-modal contextual linkages in order to obtain even better performance in describing the origins of memes.
- MIME currently focuses on single meme explanation. Exploring possibilities for generating multi-sentence explanations would improve the usability.
- The system could be developed for multi-lingual meme explanations.
- Dynamic meme capturing to enhance the dataset as memes change and go trending overnight.
- This work can be further extended and linked with emotion detection and hate-content analysis of the meme to expand its multi-domain applicability.

5. Major Strengths of the paper

- Strength-1: Multimodal approach via knowledge enriched contextualization.
- Strength-2: Dynamic meme context extraction.
- Strength-3: Empirical evaluation of model’s performance.

6. Major Weaknesses of the paper

- Weakness-1: Limitations in modeling complex abstractions.
- Weakness-2: Challenges in detecting cultural references, humour, or context-specific nuances that may not always be universally understood.
- Weakness-3: Dataset coverage and rate of data capture

7. Improvements to the paper

- Improvement-1: The paper relied on the quantitative metrics like F1 score to measure the accuracy of the proposed system. Qualitative metrics could help to evaluate specific strengths and weaknesses of MIME’s explanations. This could use human assessment of the memes and checking their coherence, relevance, and completeness with the MIME’s interpretations. This would help to understand the MIME’s performance against human judgement.

- Improvement-2: The background explanations of a meme are based on the context provided to the MIME. The provided context is input to the framework in a text file. As stated in the paper, the context is either a single or multi-line statement that potentially explains the meme. The context surrounding the meme (like source platform, hashtags, location, person tagging, comments etc.) can improve the explanations generated by MIME.