Paper Reading Task-Analyzing Hateful memes

Saideekshith Vaddineni(2022101110)

1 Summary

1.1 Importance of Memes

Memes are acknowledged as powerful tools for communication on social media platforms due to their ability to evolve across various topics such as politics, history, and sociocultural phenomena. Their adaptability makes them an ideal communication vehicle. Understanding the subtle messages conveyed within memes requires background knowledge that facilitates holistic assimilation.

1.2 Introduction

MEMEX is introduced as a new task to help understand memes better. It aims to find sentences from related contexts that can explain memes, recognizing the difficulty faced by existing multimodal models in grasping memes due to their dependence on context.

To bridge this gap, MEMEX is formed as an "evidence detection" task, which is crucial for content moderation and understanding the social impact of memes. The MCC dataset is created to support research in this area, containing both memes and their associated contextual documents.

Addressing the challenges, the MIME framework is developed, surpassing existing benchmarks. This framework enhances the MEMEX task and validates its effectiveness through empirical testing, highlighting its significant contributions.

1.3 Related Work

1.3.1 Meme Analysis

Recent studies have been conducted on understanding memes, focusing on emotions, sentiments, and themes like heroes and villains. Researchers have worked on tasks like troll meme classification, sentiment analysis, and detecting harmful content in memes using models like Visual BERT and UNITER.

1.3.2 Visual Question Answering (VQA)

The early work on VQA involves answering questions about images. Models like UpDn and LXMERT use advanced techniques like non-linear transformations and Transformers for VQA tasks.

1.3.3 Cross-modal Association

There has been an increased attention on cross-modal association due to the rise of multimodal data. It discusses methods for connecting visual concepts with text, aligning different modalities, and learning semantic relationships between visuals and text.

1.4 Meme Context Corpus (MCC)

The "Meme Context Corpus" (MCC) dataset addresses the lack of large-scale datasets capturing memes and their context. Building the MCC dataset involved meme collection, context curation, and dataset annotation.

1.4.1 Meme Collection

The dataset focuses on political, historical, and entertainment memes in English. Memes were gathered from sources like Google Images and Reddit for their diverse content.

1.4.2 Context Document Curation

Contextual documents provide background information and evidence related to memes, aiding in their interpretation. These documents support understanding meme content and themes.

1.4.3 Dataset Annotation

The dataset includes annotated evidence sentences explaining memes. Each sample comprises a meme image, context document, OCR-extracted text, and annotated evidence sentences.

1.4.4 Dataset Description

MCC covers topics like History, Entertainment, and Political Figures. Context length was limited for tractability, with an average of 128 tokens per document. The dataset was split into train, validation, and test sets for model training and evaluation.

1.5 Methodology

The methodology used is the model MIME which aims to identify explanatory evidence for memes from their related contexts. MIME takes a meme image and context as inputs and outputs labels indicating if the context's evidence explains the meme.The components of the MIME model can be summarized as follows:

1. **Knowledge-enriched Meme Encoder (KME):**
   - Incorporates external knowledge cues into meme representations.
   - Enhances the model's understanding of memes by leveraging additional information.
2. **Meme-Aware Transformer (MAT) Encoder:**
   - Specifically designed to encode meme-aware contextual information.
   - Captures the nuanced relationships between memes and their contexts.
3. **Meme-Aware LSTM (MA-LSTM):**
   - Utilizes Long Short-Term Memory (LSTM) units to capture sequential dependencies in meme-context interactions.
   - Enhances the model's ability to understand the temporal aspects of memes and their related contexts.
4. **BiLSTM Layer (BiL):**
   - Introduces a Bidirectional LSTM layer to capture contextual dependencies in both forward and backward directions.
   - Improves the model's contextual understanding by considering the entire context sequence.

The MIME framework brings together these parts to understand memes better and find clues from their surroundings. It uses both text and pictures to explain memes accurately based on what's around them.

1.6 Baseline Models

The baseline models include BERT and ViT for unimodal representations, while early fusion techniques are used for multimodal representations, combining BERT and ViT representations.

1.7 Experimental Results

The results, averaged over multiple runs, are compared using standard metrics like accuracy, F1 score, precision, recall, and E-M score. Different models, including UM Bert, ViT, MM Early Fusion, CLIP, BAN, VisualBERT, MMBT, and MIME, are evaluated on a test set. MIME demonstrates significant improvements over the best baseline model, MMBT, showcasing its effectiveness in understanding and explaining memes accurately.

1.8 Conclusion

In conclusion, this study presents the MEMEX task, which focuses on finding evidence in a context to explain memes. By creating the MCC dataset and testing different systems, as well as introducing the MIME framework, which combines meme details with context, the authors of the paper have shown that their method works well. While MIME is good at understanding context, there are still areas where it can improve. Overall, this research sets the stage or marks the beginning for better understanding memes in different situations.

1.9 Additional Details

1.9.1 Implementation Details and Hyperparameter Values (A)

A.1. The models are trained using PyTorch on an NVIDIA Tesla V100 GPU.

A.2. Pre-trained weights are imported for unimodal models, while remaining weights are randomly initialized.

A.3. Manual fine-tuning is performed over five independent runs to optimize hyperparameters.

A.4. Adam optimizer and binary cross-entropy loss function are used for training all models.

1.9.2 Additional Details about MCC (B)

B.1. Memes are collected using carefully constructed search queries from Google Images and Reddit.

B.2. Web browser automation tool Selenium is used for crawling and downloading memes.

B.3. Memes are deduplicated and filtered to remove unimodal, code-mixed, and poor-quality memes.

B.4. Context documents are curated from various sources, with Wikipedia being the primary source.

B.5. An annotation platform is built for annotating the dataset, involving two annotators briefed on the task.

1.9.3 Comparing Contexts from KYM and MIME (C)

C.1. MIME aims to get information from Wikipedia articles to complement meme understanding.

C.2. KYM provides metadata about memes but may not offer sufficient insight into the meme's intended message.

C.3. MIME's approach involves cross-modal association between memes and context documents, while KYM focuses on digital archival metadata.

C.4. Information provided by KYM may not always facilitate an understanding of the meme's intended message, as it relies on human curation and may be limited in scope.

2 Strengths of the Paper

2.1 Dataset Curation

The authors of this paper found it very necessary to curate their own dataset MCC for this task and lot of emphasis ont their entire process of curation has been explained in detail. This is a very strong point of the paper because we all know a good dataset is very essential in making o good ML model.

2.2 Simple to Understand

Though there were quite a few mathematical equations in the description of the MIME model, I could understand the structure of the model without really understanding the equations. This showes that the paper is simple to understand even without a high level understanding of the mathematics behind it.

2.3 Detailed Paper

This paper not only gives a high level overview that most people can understand but also gives a detailed explanation of the model, like hyperparameter testing, for people who want to take inspiration from this and implement something along the same path.

3 Weaknesses of the Paper

3.1 Lack of links to codebases and articles

There are a lot of links to different research papers in this paper, but not many links to codebases and articles. Links to codebases and articles related to particular model or methodology will help the person understand the inner working easily.

3.2 Lack of Definitions

There are some terminologies that are complex in nature that could require some explanation but this explanation is not provided.

3.3 Scope for Improvement

Though MIME may outperform lot of baseline models in the MEMEX task, it still has limitations especially in handling complexity of memes as indicated in the paper. Therefore, some improvement can always be made.

4 Possible Improvements to the Paper

4.1 Expand Knowledge Base

By expanding knowledge base for various historical and social subjects, The MIME model can imporve its contextual understanding of the meme.

4.2 Provide links to articles related to mathematical equations

The mathematics in this paper may not be understandable by a lot of people reading it. Hence, providing links to articles related to the mathematics involved can be very useful.

4.3 More Graphical Analysis

There could have been more graphical analysis of data and results in the paper so that it would have been easier to understand and more visually appealing.