




PAPER READING TASK

COMPUTER VISION

Name: Swayam Agrawal

Roll: 2021101068



MemeX: Detecting Explanatory Evidence for Memes via Knowledge-Enriched Contextualization

Problem Statement: Given a meme and a related document, the aim is to mine the context that explains the background of the meme succinctly.

Formulate MemeX as an “*evidence detection*” task which can help deduce pieces of contextual evidence that help bridge the information abstraction gap in the meme.

Main contributions of the paper:

- *i) A novel task, MEMEX, aimed to identify explanatory evidence for memes from their related contexts.*
- *ii) A novel dataset, MCC, containing 3400 memes and related context, along with gold-standard human annotated evidence sentence-subset.*
- *iii) A novel method, MIME that uses common sense- enriched meme representation to identify evidence from the given context.*
- *iv) Empirical analysis establishing MIME’s superiority over various unimodal and multimodal base- lines, adapted for the MEMEX task.*

The paper then describes the above contributions one by one explaining the pipeline of their implemented methods, their choices while building it and the rationale behind it.

1) MCC (Meme Context Corpus) :

- Developed in three stages: Meme Collection, Content Document Curation, and Dataset Annotation.
- The focus is on political and historical English-language memes. The reason is the higher presence of online memes based on these topics. Contextual information is curated from Wikipedia and other sources to provide background for the memes. The annotation process involves two professional annotators identifying "evidence sentences" in the context document.
- The dataset is distributed across various topics, with a significant portion dedicated to history. The dataset is split into train, validation, and test sets, each containing meme images, context documents, OCR-extracted meme text, and ground truth evidence sentences. The annotation quality is assessed using Cohen's Kappa, with substantial agreement.

MemeX: Detecting Explanatory Evidence for Memes via Knowledge-Enriched Contextualization

2) MIME (Multimodal Meme Explainer) :

- It takes a meme (an image with overlaid text) and a related context as inputs and outputs a sequence of labels indicating whether the context's constituting *evidence sentences*, either in part or collectively, explain the given meme or not. The model comprises of:
- MIME first uses two encoders (**text + multimodal encoder**) to encode contextual info and the meme (image + text).
- It then employs a **Knowledge-enriched Meme Encoder (KME)**: Enhance the understanding of memes by incorporating external common-sense knowledge. This is done using ConceptNet: a semantic network which is designed to help machines comprehend the meanings and semantic relations of the words and specific facts people use. Encoding done via a pre-trained MMBT model.
- **Meme-Aware Transformer (MAT)**: Serves as multi-layered contextual-enrichment pipeline - Enables the joint consideration of both the meme and context in the overall process of understanding the meme.
- **Meme-Aware LSTM (MA-LSTM)** : Sequential context processing and evidence detection. Cross entropy loss to optimize the model.

3) Experimentation and Performance Analysis:

- The paper experiments with various unimodal and multimodal encoders for systematically encoding memes and context representations to establish comparative baselines. Unimodal baselines include BERT and ViT while Multimodal baselines include Early fusion, MMBT, CLIP, BAN, VisualBERT.
- The metrics used for results and comparison are accuracy, F1 score (macro), precision, recall and exact-match (E-M) score. The paper observes that unimodal systems perform with mediocrity and that multimodal models either strongly compete or outperform unimodal ones (CLIP being an exception). This is for establishing baselines (MMBT performs the best with E-M score of 0.505).
- **The MIME model which is developed is noted to perform better over the best baseline established with improvements of more than 2% in each metric.** E-M score improvement is 8% : MIME E-M score is 0.585. The paper then also analyzes the detected evidences to compare the quality of detection. It is observed that the evidence predicted by MMBT does not fully explain the meme whereas those predicted by MIME are often more fitting.
- Next, the paper describes the ablation study to compare the importance of each component within the MIME model by noting the performance after removing the component from the model and replacing it by a standard transformer based component.

MemeX: Detecting Explanatory Evidence for Memes via Knowledge-Enriched Contextualization

4) Limitations:

- The image besides describes **three scenarios of ineffective detection: a) no predictions b) partial match and c) incorrect predictions.**
 - The paper next discusses the limitations in their approach - MIME. The key challenges stem from the limitations in modelling the complex level of abstractions that a meme exhibits.
- Some of the cases which the paper describes are:
1. A critical, yet a cryptic piece of information within memes, comes from the visuals, which typically requires some systematic integration of factual knowledge, that currently lacks in MIME.
 2. Insufficient textual data in the meme poses challenge for the MIME model as it will not be able to learn the required contextual associativity of the image.
 3. Risk of the model identifying incorrect or misleading evidence due to lexical bias in language used within the related context.

Meme	Related Context
	Heart of Darkness (1899) is a novella by Polish-English novelist Joseph Conrad. It tells the story of Charles Marlow, a sailor who takes on an assignment from a Belgian trading company as a ferry-boat captain in the African interior. The novel is widely regarded as a critique of European colonial rule in Africa, whilst also examining the themes of power dynamics and morality. Although Conrad does not name the river where the narrative takes place, at the time of writing the Congo Free State, the location of the large and economically important Congo River, was a private colony of Belgium's King Leopold II.
	The Jimmy Carter Peanut Statue is a monument located in Plains, Georgia, United States. Built in 1976, the roadside attraction depicts a large peanut with a toothy grin, and was built to support Jimmy Carter during the 1976 United States presidential election. The statue was commissioned by the Indiana Democratic Party during the 1976 United States presidential election as a form of support for Democratic candidate Jimmy Carter's campaign through that state. The statue, a 13-foot (4.0 m) peanut, references Carter's previous career as a peanut farmer.
	On February 26, 1815, Napoleon managed to sneak past his guards and somehow escape from Elba, slip past interception by a British ship, and return to France. Immediately, people and troops began to rally to the returned Emperor. French police forces were sent to arrest him, but upon arriving in his presence, they kneeled before him. Triumphantly, Napoleon returned to Paris on March 20, 1815. Paris welcomed him with celebration, and Louis XVIII, the new king, fled to Belgium. With Louis only just gone, Napoleon moved back into the Tuileries. The period known as the Hundred Days had begun.

Prediction errors from MIME on three *test-set* samples. The emboldened sentences in blue indicate **ground-truth evidences** and the highlighted sentences indicate model prediction

MemeX: Detecting Explanatory Evidence for Memes via Knowledge-Enriched Contextualization

Critical Analysis of the paper

1. Strengths:

- The paper addresses a relevant and interesting problem in the field of meme understanding, focusing on detecting explanatory evidence for memes. This aligns with the growing importance of analyzing and interpreting visual content in online communication.
- The development of the MemeX model is logically presented, outlining a clear methodology for meme understanding. The paper provides a structured and systematic approach to solving the problem, enhancing the model's interpretability. All components of the MIME model: Multimodal encoder, KME, MAT, MA-LSTM are explained clearly individually with a well defined architecture.
- The paper also documents the experimentations done clearly by describing the establishment of comparative baselines (unimodal + multimodal) and also excels in comparing its proposed model with existing baselines, allowing readers to realize the effectiveness of MIME against established approaches. This benchmarking provides context and demonstrates the advancements achieved by the proposed solution.
- The inclusion of an ablation study strengthens the paper by conducting a detailed analysis of the individual components of the proposed model, MIME. This showcases the significance of each component within the MIME model and its effectiveness in improving the overall model performance. The authors also provide the limitations currently present in their approach by acknowledging the challenges in modelling the complex abstractions present in memes. This also lays the ground for future work in the study of this field.

2. Weaknesses:

- The author does not cover the performance of the model over memes from other domains or languages. The dataset used focusses on political and historical English language memes. The author provides a vague statement on the same in the paper: *"The fact that such a pipeline is not constrained by a particular topic, domain, and information source makes it reasonably scalable."* There is no statistically backed data to support this claim either in the paper or in the repository associated with the paper.
- One of the other fundamental flaws lie in the fact that the paper use Wikipedia as a primary source for context document curation. The model assumes the availability of complete and accurate context from sources like Wikipedia. In real-world scenarios, this assumption may not hold, and incomplete or inaccurate context would heavily affect the model's ability to understand and interpret memes correctly.
- The paper provides no idea of the specifics of ConceptNet integration with their model: MIME, leaving room for ambiguity regarding the influence of external knowledge on the model's predictions. The repository for the code is not at documented with no details regarding the sub components of the model.

MemeX: Detecting Explanatory Evidence for Memes via Knowledge-Enriched Contextualization

Critical Analysis of the paper

3. Suggested Improvements:

- **Diversified dataset:** Expanding the dataset by incorporating memes from diverse sources beyond Google Images and Reddit. This would enhance the model's robustness and ensure a more comprehensive representation of various meme genres and cultural contexts.
- Extending the study to include memes in multiple languages to assess the model's generalizability across different linguistic contexts and subsequently an analysis of the model performance over multiple languages.
- The paper does not provide full clarity over how exactly are the standard metrics appropriate for evaluating the performance of the model, that can be provided. Another addition which can be done is to analyze and explore other metrics specifically tailored to meme understanding. Metrics that capture humour, context-awareness or semantic relevance can provide a better idea of the developed model and also provide a better analysis of the improvements in the model over baselines.
- The paper can provide a more detailed explanation of how ConceptNet is integrated into the model. This would enhance transparency and clarity regarding the influence of external common-sense knowledge on the model's decision-making process. The repository should be documented properly with proper comments in the code to indicate clearly the presence of each sub-component in the model architecture in code for ensuring no ambiguities.
- The paper should acknowledge the fact that contextual information captured from Wikipedia might not always be accurate since it is one of the fundamental assumptions in their MemeX approach and thus also should subsequently describe about the model performance and evaluation in cases when the assumption does not hold true.