Selected Papers of #AoIR2023:
The 24th Annual Conference of the
Association of Internet Researchers
Philadelphia, PA, USA / 18-21 Oct 2023

# MEMES, MULTIMODALITIES, AND MACHINES: ASSEMBLING MULTIMODAL PATTERNS IN MEME CLASSIFICATION STUDY

Guangnan Zhu
Queensland University of Technology

Kunal Chand
Queensland University of Technology

Daniel Angus
Queensland University of Technology

Timothy Graham
Queensland University of Technology

## Abstract

Memes are an important facet of current online communication on social media, with rich social, political, and cultural significance and power. This present work focuses on developing computational frameworks to support textual and visual content analysis of online memes, assisting the profiling of the unique contents and interrelationships of different meme characteristics. The framework focuses on decomposing the multimodal subcomponents of online memes to support accurate sorting and classification of meme exploitable and other rich textual materials. We showcase the development of the multimodal meme classification toolbox through a series of case studies of specific meme subcultures, with a view towards bolstering and extending existing meme analysis methods for cultural and media studies.

## Introduction

Multimodality is “the normal state of human communication” in contemporary communication (Bateman, 2014; Kress & Van Leeuwen, 2020) and particularly in online social media (Kiela et al., 2018). One of the most interesting and prevalent multimodal

Suggested Citation (APA): Zhu, G., Chand, K., Angus, D., Graham, T. (2023, October). *Memes, multimodalities, and machines: Assembling multimodal patterns in meme classification study*. Paper presented at AoIR2023: The 24th Annual Conference of the Association of Internet Researchers. Philadelphia, PA, USA: AoIR. Retrieved from http://spir.aoir.org.

information forms online is visual memes such as image macros, which are media objects made up of photos and superimposed texts, symbols and other exploitables.

Memes are popular within everyday online discourse as they allow users to encode sentiments and opinions in an engaging format, which attract people's attention, propagating to more people in little time (Johnson, 2007; Shifman, 2013). These characteristics are in line with the context of the Internet, especially contemporary political expression (Heiskanen, 2017) or as cultural capital (Nissenbaum and Shifman, 2017). Increasing evidence, however, has shown that while there are emotional expressions through this multimodality such as emotions (Nissenbaum and Shifman, 2018), memes nowadays are deeply aligned with everyday online communication that tends towards disruptive (racist, gendered, hateful content, etc.) speech and are used to convey and spread false information. For instance, the many disinformation campaigns that centred on the Pepe the Frog meme during the 2016 US presidential election. Thus, there is a need for more complete and reliable methods and tools for fine-grained but large-scale meme analysis.

Computer vision scholars have previously deployed computer vision techniques to study the visual patterns of memes posted on social media (He et al., 2016). Recent studies also deploy Optical Character Recognition (OCR) to extract text from images, utilising these different modalities through advanced computer vision and natural language processing models (Du et al., 2020; Singh et al., 2021). Some studies also propose streaming pipeline frameworks to detect and compare trending events with online meme generator websites (e.g., Know Your Meme) to cluster images into meme template groups (JafariAsbagh et al., 2014; Zannettou et al., 2018).

Despite many breakthroughs in the development of computational methods for detecting and classifying memes, these techniques are still limited in how well they support exploratory analysis of online memes. This limitation is largely due to the multimodality of memes, particularly the intersection of these modalities where these components play a joint role in conveying the meaning of the meme (Smitha et al., 2018). Furthermore, previous methods often work as end-to-end 'black box' systems that prevent researchers from interrogating and understanding how and why specific inputs are classified or clustered in specific ways (Burgess, et al., 2021). These methods may also ignore the latent political and cultural patterns reflected in isolated modalities while some other qualitative analysis methods are hard to deploy on large-scale meme detection and analysis (Nissenbaum and Shifman, 2018).

Owing to these critical gaps, the present work focuses on developing a computational framework that can separate different components of image macro memes into their constituent multimodal features for examination while also supporting combinations of modalities where such combinations have direct consequences for interpretations of the memetic material. By doing so this method contributes a more complete and extensible meme study framework with conceptual and methodological approaches for media and cultural studies scholars.

**Meme Machine: A novel multimodal Computational Framework for Meme analysis**

The Meme Machine is a computational meme analysis framework that provides 3 main functionalities:

- classify memes based on existing known exploitable or other signifiers;
- cluster memes together based on visual similarity; and,
- cluster memes together based on semantic similarity.

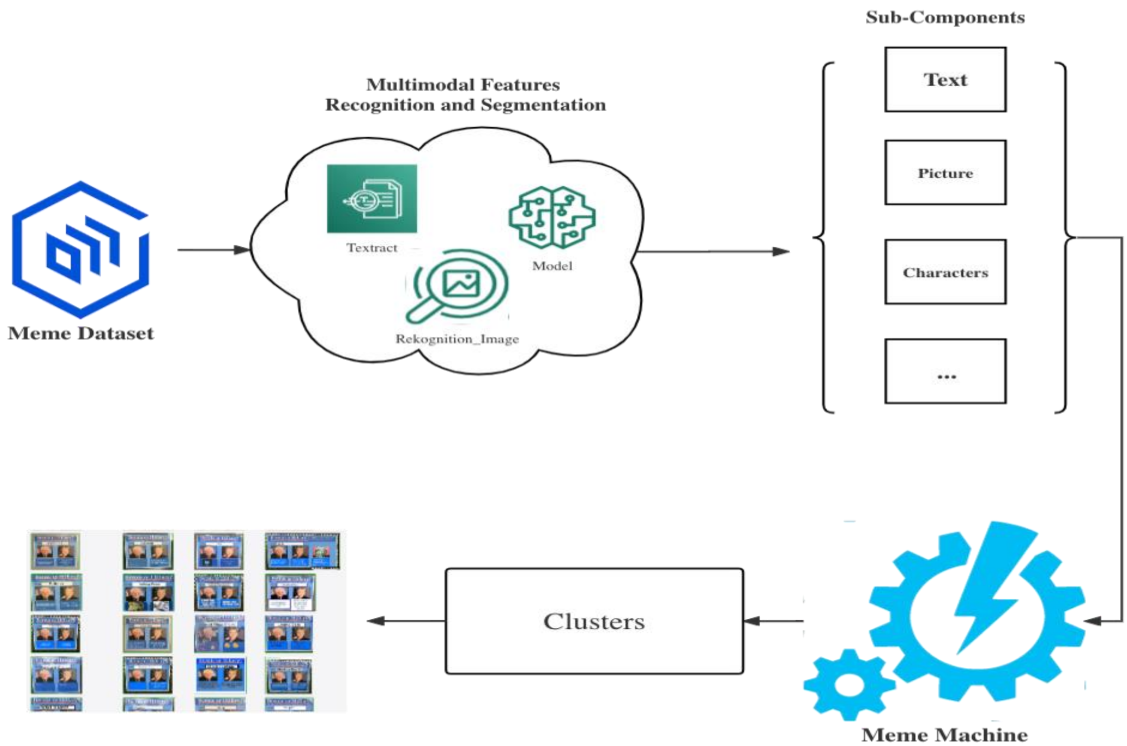


**Figure 1**. *The Meme Machine integrates various multimodal algorithms and models to extract different features, segmenting the original image into subcomponents. The subcomponents are fed further into meme machine to obtain clusters of memes based on the subcomponents.*

The first step in the meme machine is to decompose the input meme into its various constituent parts. This decomposition is achieved using a combination of text detection methods and a single-shot classification approach pre-trained on known exploitable materials, sourced from common online meme dictionaries and taxonomies. This decomposition process ties in with the first functionality in providing confidence scores for the presence of specific exploitable within the input.

Clustering through visual similarity operates by running decomposed images through a pretrained image classification model to extract feature vectors, used to group other images with similar feature vector values. Semantic similarity works in a similar manner except that text is run through an additional pretrained topic model to extract feature vectors.

To enable these processes, we use a variety of existing machine learning approaches, stitched together to help identify components of memes. An example of one approach deployed is Tesseract, an OCR engine that can detect text within an image and provide locations of where text was detected. Based on these locations, those multimodal components can be segmented and treated independently of each other.

We also make use of the ‘Image Machine’, a machine vision tool based on the critical simulation methodology that was designed to help scholars understand how machine learning technologies operate (Burgess, et al., 2021). The image machine provides functions that allow the unsupervised clustering of sets of input images into emergent clusters based on shared visual features and latent aesthetic features. After feeding the decomposed visual components through the tool, we receive clustering information of similar components. This information can be combined with other clustering data to assist in the overall clustering and organisation of memetic materials for interpretative analysis.

**Result**

While this toolkit is still a work in progress, initial experiments in feeding a dataset of memes through the tool show that it can isolate various themes and patterns from the dataset of memes. An example is seen below, where the original input of memes is seen compared to the output structured set of memes after passing it through the tool.

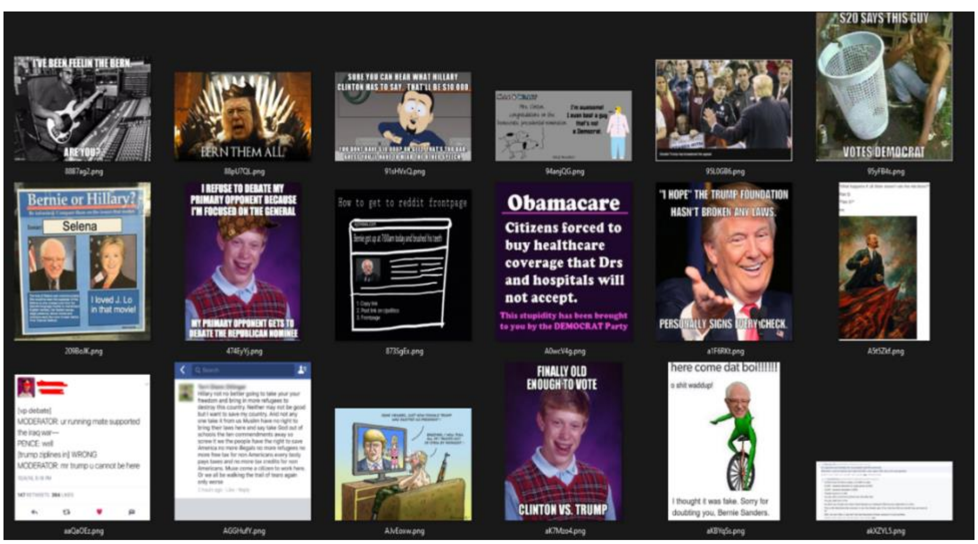


**Figure 2.** *The sample of original input from the total meme dataset before passing through the tool.*

In this image, we can see a small portion of the total meme dataset, including a variety of styles and messages within the memes, all of which is unstructured (not organised by any metric except name). By putting it through the tool, this collection of images will be structured such that visually similar images (memes) will be clustered together.

In Figure 4 we see what happens when the last cluster (lower right in Figure 3) is selected. Memes within this cluster (group) can be seen to be highly visually similar, sharing a header with a red underline, a subheader whose text varies, and side-by-side portraits of Democratic Presidential nominees Bernie Sanders and Hillary Clinton, each portrait having various text underneath.

The Meme Machine example above is achieved without extracting text from the images, so these memes were not grouped by anything text related. This capability for semantic clustering is being explored currently, and will be added in upcoming versions including text-based clustering (such as topic modelling and sentiment analysis) as additional inputs into the meme cluster determination.

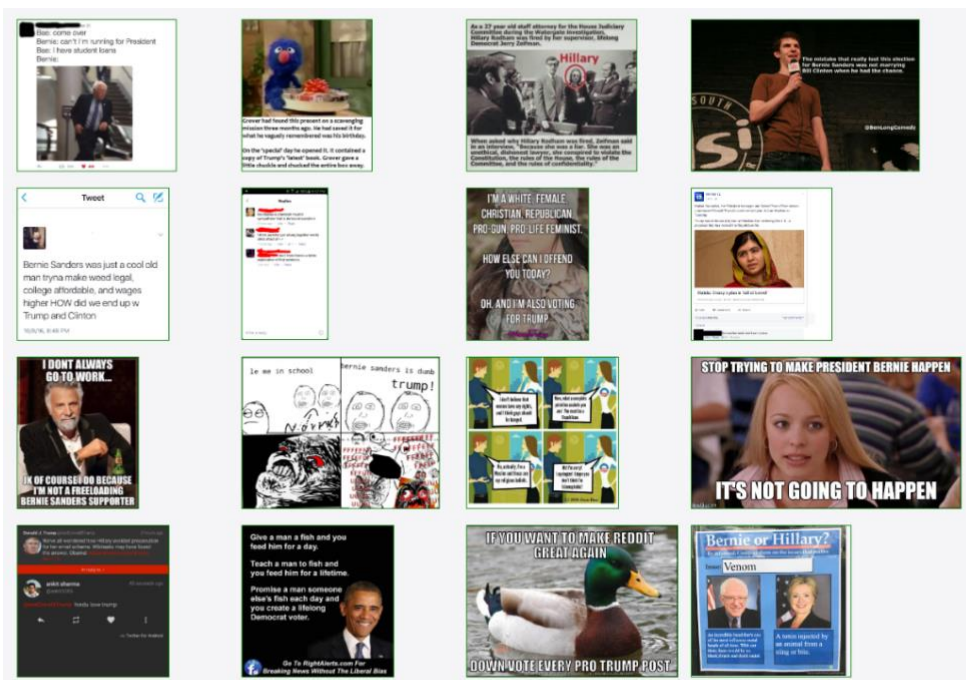


**Figure 3.** *16 groups of memes are identified here, each with a centroid image of each cluster as a summary of memes located in this cluster. These meme clusters have distinct visual similarities in their use of exploitables and overall composition.*


**Figure 4**. *The memes within the same clusters after passing through the tool have a strong visual similarity. However, present progress hasn't conducted the similarity reflected in such as semantic patterns or other modalities.*

## References

Bateman, J. A. (2014). Text and image: A critical introduction to the visual/verbal divide. Routledge.

Burgess, J., Angus, D., Carah, N., Andrejevic, M., Hawker, K., Lewis, K., … Li, L. (2021, November 6). *Critical simulation as hybrid digital method for exploring the data operations and vernacular cultures of visual social media platforms.* https://doi.org/10.31235/osf.io/2cwsu

Du, Y., Masood, M. A., & Joseph, K. (2020, May). Understanding visual memes: An empirical analysis of text superimposed on memes shared on twitter. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 14, pp. 153-164).

He, S., Zheng, X., Wang, J., Chang, Z., Luo, Y., and Zeng, D. (2016). Meme extraction and tracing in crisis events. In 2016 IEEE Conference on Intelligence and Security Informatics (ISI), pages 61–66, Tucson, AZ, USA. IEEE.

Heiskanen, B. (2017). Meme-ing electoral participation. European journal of American studies, 12(12-2).

JafariAsbagh, M.; Ferrara, E.; Varol, O.; Menczer, F.; and Flammini, A. 2014. Clustering memes in social media streams. Social Network Analysis and Mining 4:1–13.

Johnson, D. (2007). Mapping the meme: A geographical approach to materialist rhetorical criticism. Communication and Critical/Cultural Studies, 4(1), 27–50.

Kiela, D., Grave, E., Joulin, A., & Mikolov, T. (2018). Efficient large-scale multi-modal classification. In Proceedings of the aaai conference on artificial intelligence (Vol. 32).

Kress, G., & Van Leeuwen, T. (2020). Reading images: The grammar of visual design. Routledge.

Nissenbaum, A., & Shifman, L. (2017). Internet memes as contested cultural capital: The case of 4chan's/b/board. New media & society, 19(4), 483-501.

Nissenbaum, A., & Shifman, L. (2018). Meme templates as expressive repertoires in a globalizing world: A cross-linguistic study. Journal of Computer-Mediated Communication, 23(5), 294-310.

Shifman, L. (2013). Memes in digital culture. MIT press.

Singh, V. K., Ghosh, I., & Sonagara, D. (2021). Detecting fake news stories via multimodal analysis. Journal of the Association for Information Science and Technology, 72(1), 3–17.

Smitha, E. S., Sendhilkumar, S., & Mahalaksmi, G. S. (2018). Meme classification using textual and visual features. In Computational Vision and Bio Inspired Computing (pp. 1015-1031). Springer International Publishing.

Zannettou, S.; Caulfield, T.; Blackburn, J.; De Cristofaro, E.; Siri- vianos, M.; Stringhini, G.; and Suarez-Tangil, G. 2018. On the origins of memes by means of fringe web communities. In Pro- ceedings of the Internet Measurement Conference 2018, IMC '18, 188–202. New York, NY, USA: ACM.