Multimodal Sentiment Analysis

Alejandro Ciuba, Modhumonty Das and **Nick Littlefield**
School of Computing and Information
University of Pittsburgh, Pittsburgh, PA, USA
{alejandrociuba, mod53, ngl18}@pitt.edu

1 Introduction

Sentiment analysis models user sentiments via texts, images and videos. Its applications vary from sociolinguistic research to marketing analytics (Alessia et al., 2015; Nguyen et al., 2016). Sentiment analysis tasks usually focuses on one modality (e.g. reading food reviews), but this can leave out crucial context (e.g. the food's appearance). Due to unimodality's limitations, the push for multimodal models has grown, with many believing it is the necessary next step for AI (Bender and Koller, 2020). This creates an interesting research space to examine modality contributions towards model performance. We plan to study several smaller-scale unimodal and multimodal sentiment analysis approaches. Our initial research questions are: **(RQ1)** How do unimodal methods compare when given only text/image input? **(RQ2)** How do these methods, when adapted to a multimodal setting, compare with their unimodal counterparts? **(RQ3)** How does multimodality influence our methods? **(RQ4)** And can we potentially compare our approaches with state-of-the-art multimodal models?

2 Related Work

Existing work emphasizes the importance of multiple modalities for better affective analysis[1] (Shoumy et al., 2020). Achlioptas et al., 2021 created the ArtEmis dataset to develop models for emotion prediction given images and texts. A novel, deep neural network approach to multimodal sentiment analysis was explored in Hu and Flaxman, 2018. Their goal was to infer latent emotional user states via "emotion word tags" attached to Tumblr posts. The model was validated on an image set used in psychological studies, and the combined multimodal information was used for the tag prediction task. The effectiveness of multimodal analysis has also been evaluated by comparing computational emotion classification approaches applied to face videos and bio-sensing modalities in Siddharth et al., 2018, which addressed the limitations of single-sensing modalities, and showed the advantages and increased accuracy of multimodal affective computing.

3 Dataset & Challenges

We plan to leverage the ArtEmis dataset[2], as it is a valuable resource for uni- and multimodal experiments due to art's subjectivity. The dataset has 80,031 paintings spanning 27 styles and 45 genres. It includes an annotation set of emotional and explanatory utterances from over 6,788 annotators. In total, the dataset has a collection of 454,684 explanatory utterances associated with nine emotional reactions (e.g. amusement, anger, disgust, fear, sadness, etc.). The first challenge is understanding the dataset itself due to its large scale and multimodality, which is a prerequisite before implementing any methods. The second is that there are many utterances (and emotions) per art piece. Lastly, the distribution of art styles is imbalanced and will need to be handled before building many of the classifiers (e.g. neural models).

4 Proposed Methodology

Three modality groups designed for the dataset's sentiment analysis task will be created: "only utterances", "only paintings" and lastly "multimodal approaches," which will utilize both the utterances and art. The results for the different classification methods across and within the modality groups will then be statistically analyzed. Basic classifiers such as SVMs, Naive Bayes', and deep neural networks can be used, as well as more advanced models like CLIP (Radford et al., 2021).

[1] Sentiment analysis beyond standard positive, negative and neutral tagging (Molina Beltrán et al., 2019).

[2] `https://www.artemisdataset.org/` is built upon the WikiArt dataset `https://www.wikiart.org/`

References

Panos Achlioptas, Maks Ovsjanikov, Kilichbek Haydarov, Mohamed Elhoseiny, and Leonidas Guibas. 2021. ArtEmis: Affective Language for Visual Art. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11564–11574, Nashville, TN, USA. IEEE.

D Alessia, Fernando Ferri, Patrizia Grifoni, and Tiziana Guzzo. 2015. Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3). Publisher: Citeseer.

Emily M. Bender and Alexander Koller. 2020. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online. Association for Computational Linguistics.

Anthony Hu and Seth Flaxman. 2018. Multimodal Sentiment Analysis To Explore the Structure of Emotions. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 350–358, London United Kingdom. ACM.

Carlos Molina Beltrán, Alejandra Andrea Segura Navarrete, Christian Vidal-Castro, Clemente Rubio-Manzano, and Claudia Martínez-Araneda. 2019. Improving the affective analysis in texts. *The Electronic Library*, 37(6):984–1006. Publisher: Emerald Publishing Limited.

Dong Nguyen, A. Seza Doğruöz, Carolyn P. Rosé, and Franciska de Jong. 2016. Computational Sociolinguistics: A Survey. *Computational Linguistics*, 42(3):537–593.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, and others. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

Nusrat J. Shoumy, Li-Minn Ang, Kah Phooi Seng, D. M. Motiur Rahaman, and Tanveer Zia. 2020. Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals. *Journal of Network and Computer Applications*, 149:102447.

Siddharth, Tzyy-Ping Jung, and Terrence J. Sejnowski. 2018. Multi-modal Approach for Affective Computing. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 291–294, Honolulu, HI. IEEE.