Multimodal Data Fusion with Quantum Inspiration

Qiuchi Li

qiuchili@dei.unipd.it

University of Padua

Padua, Italy

ABSTRACT

Language understanding is multimodal. During human communication, messages are conveyed not only by words in textual form, but also through speech patterns, gestures or facial emotions of the speakers. Therefore, it is crucial to fuse information from different modalities to achieve a joint comprehension.

With the rapid progress in the deep learning field, neural networks have emerged as the most popular approach for addressing multimodal data fusion [1, 6, 7, 12]. While these models can effectively combine multimodal features by learning from data, they nevertheless lack an explicit exhibition of how different modalities are related to each other, due to the inherent low interpretability of neural networks [2].

In the meantime, Quantum Theory (QT) has given rise to principled approaches for incorporating interactions between textual features into a holistic textual representation [3, 5, 8, 10], where the concepts of *superposition* and *entanglement* have been universally exploited to formulate interactions. The advantages of those models in capturing complicated correlations between textual features have been observed.

We hereby propose the research on quantum-inspired multimodal data fusion, claiming that the limitation of multimodal data fusion can be tackled by quantum-driven models. In particular, we propose to employ *superposition* to formulate intra-modal interactions while the interplay between different modalities is expected to be captured by entanglement measures. By doing so, the interactions within multimodal data may be rendered explicitly in a unified quantum formalism, increasing the performance and interpretability for concrete multimodal tasks. It will also expand the application domains of quantum theory to multimodal tasks where only preliminary efforts have been made [11]. We therefore aim at answering the following research question:

RQ. Can we fuse multimodal data with quantum-inspired models?

To answer this question, we propose to fuse multimodal data with complex-valued neural networks, motivated by the theoretical link between neural networks and quantum theory [4] and advances in complex-valued neural networks [9]. Our model begins with a separate complex-valued embedding learned for each unimodal data based on the existing works [5, 10], which inherently assumes *superposition* between intra-modal features. Then we construct a many-body system in entangled state for multimodal data,

where cross-modality interactions are naturally reflected by entanglement measures. Quantum measurement operators are applied to the entanglement state to address a concrete multimodal task at hand. The whole process is instrumented by a complex-valued neural network, which is able to learn how multimodal features are combined from data, and at the same time explain the combination by means of quantum superposition and entanglement measures.

We plan to examine our proposed models on CMU-MOSI [12] and CMU-MOSEI [1], which are benchmarking multimodal sentiment analysis datasets. The dataset targets at classifying sentiment into 2, 5 or 7 classes with the input of textual, visual and acoustic features. We expect to see comparable effectiveness to state-of-the-art models, and we will explore superposition and entanglement measures to better understand the inter-modal interactions.

CCS CONCEPTS

• **Information systems** → **Multimedia content creation**; *Document representation*; *Content analysis and feature selection*; *Sentiment analysis*.

KEYWORDS

Multimodal data fusion, Quantum theory, Neural network

REFERENCES

[1] Amirali Bagher Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph. In *ACL 2018*.

[2] Tadas Baltrusaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal Machine Learning: A Survey and Taxonomy. *arXiv:1705.09406 [cs]*, 2017.

[3] William Blacoe. On Quantum Generalizations of Information-Theoretic Measures and their Contribution to Distributional Semantics. *arXiv:1506.00578 [cs]*, 2015.

[4] Yoav Levine, David Yakira, Nadav Cohen, and Amnon Shashua. Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design. In *ICLR*, 2018.

[5] Qiuchi Li, Benyou Wang, and Massimo Melucci. CNM: An Interpretable Complex-valued Network for Matching. *arXiv:1904.05298 [cs]*, April 2019.

[6] Soujanya Poria, Erik Cambria, and Alexander Gelbukh. Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-level Multimodal Sentiment Analysis. In *EMNLP 2015*.

[7] Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. Context-Dependent Sentiment Analysis in User-Generated Videos. In *ACL 2017*.

[8] Alessandro Sordoni, Jing He, and Jian-Yun Nie. Modeling Latent Topic Interactions Using Quantum Interference for Information Retrieval. In *CIKM 2013*.

[9] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J. Pal. Deep Complex Networks. February 2018.

[10] Benyou Wang, Qiuchi Li, Massimo Melucci, and Dawei Song. Semantic Hilbert Space for Text Representation Learning. *The Web Conference 2019*.

[11] Jun Wang, Dawei Song, and Leszek Kaliciak. Tensor product of correlated text and visual features. In *QI 2010*.

[12] Amir Zadeh, Rown Zellers, and Eli Pincus. MOSI: Multimodal Corpus of Sentiment Intensity and Subjectivity Analysis in Online Opinion Videos. *IEEE Intelligent Systems*, 2016.