

Multimodal Emotion-Cause Pair Extraction in Conversations

Konari Yuvaraju, Katchi Vinay, Lingala Samhas

Department of Computer Science and Engineering
R.V.R. & J.C. College of Engineering, Guntur, India

Abstract: *In this work, we present a modular pipeline for Emotion-Cause Pair Extraction (ECPE) in conversational data, designed to operate across multi-modal sources such as text, audio transcripts, and video subtitles. Unlike traditional systems that rely on annotated datasets and end-to-end training, our approach leverages pre-trained models for emotion classification and cause inference to extract meaningful emotion-cause pairs directly from real-world dialogue. We integrate a BERT-based emotion classifier with a question-answering model to identify both the emotion expressed in an utterance and the underlying cause from the surrounding context. This framework enables researchers and developers to analyze emotions and their triggers without the overhead of dataset creation or domain-specific fine-tuning. While we do not perform direct video annotation, our system supports scalable post-hoc analysis, making it useful as a semi-automated toolkit for annotating conversational datasets. The novelty of our work lies in fusing independent inference models into a unified ECPE pipeline that can be extended to support annotation, research, or downstream dialogue applications. Our method provides a practical step toward real-time ECPE inference in resource-constrained and low-data environments*

Keywords: Emotion-Cause Pair Extraction (ECPE), Multi-modal Emotion Analysis, Pre-trained Transformers (BERT, RoBERTa), Real-time Conversational Inference

