VerbaNexAI Lab at SemEval-2024 Task 3: Deciphering emotional causality in conversations using multimodal analysis approach

Victor Pacheco and **Juan Cuadrado** and **Elizabeth Martinez**
and **Juan Carlos Martinez-Santos** and **Edwin Puertas**
Instituto Tecnologico de Tepic, Tepic México
Universidad Tecnologica de Bolivar, Cartagena Colombia
epuerta@utb.edu.co

Abstract

This study delineates our participation in the SemEval-2024 Task 3: Multimodal Emotion Cause Analysis in Conversations, focusing on developing and applying an innovative methodology for emotion detection and cause analysis in conversational contexts. Leveraging logistic regression, we analyzed conversational utterances to identify emotions per utterance. Subsequently, we employed a dependency analysis pipeline, utilizing SpaCy to extract significant chunk features, including object, subject, adjectival modifiers, and adverbial clause modifiers. These features were analyzed within a graph-like framework, conceptualizing the dependency relationships as edges connecting emotional causes (tails) to their corresponding emotions (heads). Despite the novelty of our approach, the preliminary results were unexpectedly humbling, with a consistent score of 0.0 across all evaluated metrics. This paper presents our methodology, the challenges encountered, and an analysis of the potential factors contributing to these outcomes, offering insights into the complexities of emotion-cause analysis in multimodal conversational data.

1 Introduction

The automatic detection and analysis of emotions in the text have gained substantial traction within the realm of computational linguistics and affective computing, particularly given their profound implications for understanding human-computer interaction (Moreno-Sandoval et al., 2019; Puertas et al., 2021), mental health, and the dynamics of social communication (Cuadrado et al., 2023a; Puertas et al., 2022). Integrating emotion analysis technologies into interactive systems can revolutionize user experience, providing applications that can adapt to users' emotional states in real-time, thus creating more engaging and personalized interactions (Moreno-Sandoval et al., 2020). SemEval-2024 Task 3 introduces a compelling challenge in this arena: the Multimodal Emotion Cause Analysis in Conversations (Wang et al., 2024). This task necessitates the identification of emotions within conversational utterances and elucidating the underlying causes behind these emotional expressions, a nuanced inquiry that extends beyond the textual modality to encompass a multimodal understanding.

In mental health, the automated text analysis for emotional content has shown promise as a tool for early detection of disorders such as depression and anxiety, using data from social media platforms as a rich source of behavioral indicators (Martinez et al., 2023). Similarly, in the domain of social communication, understanding the interplay of emotions and their causes in conversations can enhance the development of systems designed for conflict resolution, educational purposes, and even in the entertainment industry to create more compelling narratives and interactive experiences (Cuadrado et al., 2023b; Moreno-Sandoval et al., 2020; Marrugo-Tobón et al., 2023; Puertas and Martinez-Santos, 2021; Puertas et al., 2019).

Motivated by this task's complexity and innovative potential, our study endeavors to bridge the gap between emotion detection and cause analysis through a methodological approach. By integrating logistic regression models for emotion identification with advanced dependency analysis techniques for cause extraction, we aimed to uncover the intricate patterns of emotional causality in conversations. Our approach is distinguished by its reliance on linguistic features extracted from dependency graphs, facilitating a deeper understanding of the relational structures that underpin emotional discourse.

However, the journey from conceptualization to realization was fraught with challenges. Initial results, marked by a uniform score of 0.0 across evaluation metrics, starkly underscored the task's difficulty and highlighted our methodology's limita-

Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pages 1339–1343

tions. This outcome reflects the complexities inherent in the nuanced analysis of emotional causality. It emphasizes the need for continuous innovation in the field. The exploration of multimodal data sources, including text, audio, and visual cues, represents a frontier in affective computing, promising to enrich our understanding of emotions in human interactions.

2 Related Work

The task of emotion-cause analysis in conversations has been extensively explored, with recent advancements highlighting the integration of multimodal data and sophisticated natural language processing techniques. This section synthesizes contributions from several seminal works that are directly relevant to the challenges and objectives of SemEval-2024 Task 3 (Wang et al., 2024, 2023; Xia and Ding, 2019).

The MELD dataset introduced by Poria et al. presents a comprehensive multimodal multi-party dataset for emotion recognition in conversations, underscoring the significance of leveraging verbal and non-verbal cues for accurate emotion detection (Poria et al., 2018). This work emphasizes the complexity of emotion and sarcasm recognition in dynamic conversational contexts. It sets a foundational benchmark for subsequent research.

Kumar et al.'s exploration of emotion and its flip in multi-party conversations through a masked memory network and transformer model offers profound insights into the transitional dynamics of emotional states within dialogue (Kumar et al., 2022). Their methodology provides a robust framework for understanding emotional shifts, contributing valuable perspectives to conversational emotion analysis.

Research on multimodal sarcasm detection and humor classification in code-mixed conversations by (Christ et al., 2023) illustrates the challenges and opportunities in identifying nuanced affective states through a combination of textual and visual information. This study highlights the intricate relationship between language use and non-verbal indicators in conveying sarcasm and humor.

The work by (Gupta et al., 2023) on explaining sarcastic utterances to enhance affect understanding in multimodal dialogues further delves into the importance of context and explicit presentation of utterances for sarcasm detection. Their approach underscores the necessity of comprehensive analysis to grasp the underlying affective dimensions in conversations.

The investigation into the reasoning capabilities of large language models by (Xu et al., 2023) questions their efficacy in logical deduction and emotional understanding within conversational settings. This critique prompts a reevaluation of the applications of such models in affective computing, suggesting a need for more nuanced and context-aware methodologies.

Recent studies introduce novel problems and solutions, such as sarcasm explanation in multimodal multi-party dialogues (Yadav et al., 2021), targeted sentiment analysis in the financial domain using transformer-based models and phonestheme embeddings (Vasanth et al., 2022), and automated depression detection leveraging lexical features and RoBERTa transformer models (Naseem et al., 2020). These contributions highlight the expanding scope of affective computing research, including advancements in hate speech detection (Sisi and Jing, 2023) and the nuanced analysis of economic sentiment (Vasanth et al., 2022), underscoring the versatility and complexity of emotion analysis in diverse contexts.

3 Methodology

We designed our methodology to address the challenge of emotion-cause analysis in Conversations, as outlined in SemEval-2024 Task 3, focusing exclusively on textual analysis before considering other modalities. This decision stemmed from our initial goal to establish a solid foundation in text analysis, which, if successful, would have been extended to a multimodal approach. Unfortunately, our endeavors in the textual dimension did not yield the expected results, leading us to concentrate on addressing these challenges without extending our analysis to include audio or video data. We made accessible all scripts and data related to this study at SemEval 2024 VerbaNex AI Repository.

3.1 Data Acquisition and Preprocessing

Our initial phase involved collecting a dataset of conversational utterances provided by the SemEval-2024 Task 3 organizers. Despite the task's multimodal nature, we primarily focused on the textual component. We aimed to extract and analyze linguistic features indicative of emotional expression and causality using standard NLP techniques such as tokenization, lemmatization, and part-of-speech

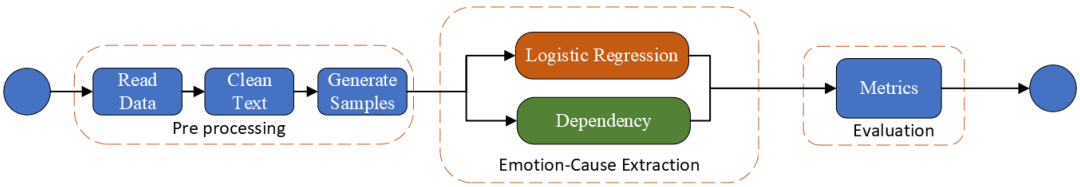

Figure 1: System General Pipeline

tagging facilitated by the SpaCy NLP library.

3.2 Emotion Detection Using Logistic Regression

We implemented a logistic regression model for emotion detection, chosen for its simplicity and effectiveness in binary classification tasks. We trained this model on a dataset subset labeled with emotional annotations. We based the features selected for the model on their potential relevance to emotional expression, including word embeddings derived from pre-trained models and syntactic features extracted during preprocessing. Additionally, we utilized 'CountVectorizer' from scikit-learn to transform text documents into a numerical matrix of word or token counts, which aided in the feature extraction process.

3.3 Dependency Analysis and Feature Extraction

Our core methodology involved analyzing dependency structures within conversational utterances. Utilizing SpaCy's dependency parser and basing our approach on Universal Dependency Relations and the concepts outlined in the "Universal Stanford Dependencies: A cross-linguistic typology," we identified and analyzed noun chunks and their syntactic relationships. We focused on dependencies indicative of emotional causality, such as objects, subjects, adjectival modifiers, and adverbial clause modifiers. These dependency relationships were conceptualized as edges in a graph-like structure, aiding in analyzing the connection between emotional expressions and their causes within the conversation.

3.4 Analysis for Emotion Cause Identification

The culmination of our methodology involved applying a graph-based analysis to map the extracted dependency features onto a framework modeling the conversational flow and interconnections between utterances. This analysis aimed to identify patterns signifying the causality of emotions by tracing the path from an emotional expression to its trigger within the graph. Despite the theoretical soundness of this approach, it encountered challenges in accurately capturing the nuances of emotional causality.

3.5 Task Formulation and Model Inputs

The task was formulated to detect emotion-cause pairs within a complete conversation, with each conversation serving as a single input to our model. We designed this approach to capture the intricacies of conversational dynamics and the contextual interplay of emotional expressions and their causes.

4 Partial Results and Failure Analysis

The logistic regression model achieved a score of 0.52 in emotion detection. However, the final results for emotion-cause pairing were disappointingly at 0.0. The initial low precision of the emotion detector contributed to these results. Further complicating our analysis, handling punctuation and syntactic ambiguities in dependency analysis introduced additional challenges, underscoring the complexity of accurately analyzing emotional causality in text.

5 Conclusions

Using logistic regression and dependency analysis techniques, our exploration into Emotion Cause Analysis in Conversations encountered unexpected hurdles, culminating in a consistent score of 0.0 across all evaluation metrics. This outcome reflects the inherent challenges in interpreting emotional cues from text alone. It highlights a fundamental mismatch between our initial methodological choices and the task's complexity. Our experience underlines the intricacies of detecting and analyzing emotional expressions and their causes within conversational dynamics. Our chosen tools did not capture the nuanced interplay of emotions in a multimodal context.

6 Future Directions

Our research will pivot towards employing more sophisticated models, focusing on transformer-based approaches. These models are renowned for their ability to grasp deeper contextual nuances and hold promise for significantly improving our analysis. Recognizing the pivotal role of multimodal data in enriching emotion analysis, we aim to extend our methodology to incorporate audio and visual cues alongside textual information. This expansion might offer a more holistic understanding of emotional causality in conversations. By refining our evaluation framework and integrating these diverse data types, we aspire to contribute meaningfully to the fields of affective computing and computational linguistics, paving the way for a more nuanced and comprehensive exploration of emotional interactions in conversational settings.

Acknowledgments

To the SemEval contest, sponsored by the SIGLEX Special Interest Group on the Lexicon of the Association for Computational Linguistics. To the master's degree scholarship program in engineering at the Universidad Tecnológica de Bolivar (UTB) in Cartagena, Colombia.

We would like to express our gratitude to the team at the VerbaNex AI Lab [1] for their dedication, collaboration, and ongoing support of our research endeavors.

References

Lukas Christ, Shahin Amiriparian, Alice Baird, Alexander Kathan, Niklas Müller, Steffen Klug, Chris Gagne, Panagiotis Tzirakis, Lukas Stappen, Eva-Maria Meßner, et al. 2023. The muse 2023 multimodal sentiment analysis challenge: Mimicked emotions, cross-cultural humour, and personalisation. In *Proceedings of the 4th on Multimodal Sentiment Analysis Challenge and Workshop: Mimicked Emotions, Humour and Personalisation*, pages 1–10.

J. Cuadrado, E. Martinez, J.C. Martinez-Santos, and E. Puertas. 2023a. Team utb-nlp at finances 2023: Financial targeted sentiment analysis using a phonestheme semantic approach. *CEUR Workshop Proceedings*, 3496.

Juan Cuadrado, Elizabeth Martinez, Anderson Morillo, Daniel Peña, Kevin Sossa, Juan Martinez-Santos, and Edwin Puertas. 2023b. Utb-nlp at semeval-2023 task 3: Weirdness, lexical features for detecting categorical framings, and persuasion in online news. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1551–1557.

Rishabh Gupta, Shaily Desai, Manvi Goel, Anil Bandhakavi, Tanmoy Chakraborty, and Md Shad Akhtar. 2023. Counterspeeches up my sleeve! intent distribution learning and persistent fusion for intent-conditioned counterspeech generation. *arXiv preprint arXiv:2305.13776*.

Shivani Kumar, Anubhav Shrimal, Md Shad Akhtar, and Tanmoy Chakraborty. 2022. Discovering emotion and reasoning its flip in multi-party conversations using masked memory network and transformer. *Knowledge-Based Systems*, 240:108112.

Duván Andres Marrugo-Tobón, Juan Carlos Martinez-Santos, and Edwin Puertas. 2023. Natural language content evaluation system for multiclass detection of hate speech in tweets using transformers. *In Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023).*

E. Martinez, J. Cuadrado, D. Peña, J.C. Martinez-Santos, and E. Puertas. 2023. Automated depression detection in text data: Leveraging lexical features, phonesthemes embedding, and roberta transformer model. *CEUR Workshop Proceedings*, 3496.

Luis Gabriel Moreno-Sandoval, Edwin Puertas, Flor Miriam Plaza-del Arco, Alexandra Pomares-Quimbaya, Jorge Andres Alvarado-Valencia, and L Alfonso. 2019. Celebrity profiling on twitter using sociolinguistic. *CLEF (Working Notes)*.

Luis Gabriel Moreno-Sandoval, Edwin Puertas, Alexandra Pomares-Quimbaya, , and Jorge Andres Alvarado-Valencia. 2020. Assembly of Polarity, Emotion and User Statistics for Detection of Fake Profiles—Notebook for PAN at CLEF 2020. In *CLEF 2020 Labs and Workshops, Notebook Papers*. CEUR-WS.org.

Usman Naseem, Imran Razzak, Katarzyna Musial, and Muhammad Imran. 2020. Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Generation Computer Systems*, 113:58–69.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2018. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*.

Edwin Puertas and Juan Carlos Martinez-Santos. 2021. Phonetic detection for hate speech spreaders on twitter. In *CLEF*.

Edwin Puertas, Juan Carlos Martinez-Santos, and Pablo Andrés Pertuz-Duran. 2022. Presidential preferences in colombia through sentiment analysis. In *2022 IEEE ANDESCON*, pages 1–5.

[1] https://github.com/VerbaNexAI

Edwin Puertas, Luis Gabriel Moreno-Sandoval, Flor Miriam Plaza-del Arco, Jorge Andres Alvarado-Valencia, Alexandra Pomares-Quimbaya, and L Alfonso. 2019. Bots and gender profiling on twitter using sociolinguistic features. *CLEF (Working Notes)*, pages 1–8.

Edwin Puertas, Luis Gabriel Moreno-Sandoval, Javier Redondo, Jorge Andres Alvarado-Valencia, and Alexandra Pomares-Quimbaya. 2021. Detection of sociolinguistic features in digital social networks for the detection of communities. *Cognitive Computation*, 13:518–537.

Wu Sisi and Ma Jing. 2023. Multi-task & multi-modal sentiment analysis model based on aware fusion. *Data Analysis and Knowledge Discovery*, 7(10):74–84.

K Vasanth, P Sridevipon malar, Vikram Shete, CN Ravi, et al. 2022. Dynamic fusion of text, video and audio models for sentiment analysis. *Procedia Computer Science*, 215:211–219.

Fanfan Wang, Zixiang Ding, Rui Xia, Zhaoyu Li, and Jianfei Yu. 2023. Multimodal emotion-cause pair extraction in conversations. *IEEE Transactions on Affective Computing*, 14(3):1832–1844.

Fanfan Wang, Heqing Ma, Rui Xia, Jianfei Yu, and Erik Cambria. 2024. Semeval-2024 task 3: Multimodal emotion cause analysis in conversations. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 2022–2033, Mexico City, Mexico. Association for Computational Linguistics.

Rui Xia and Zixiang Ding. 2019. Emotion-cause pair extraction: A new task to emotion analysis in texts. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1003–1012.

Fangzhi Xu, Qika Lin, Jiawei Han, Tianzhe Zhao, Jun Liu, and Erik Cambria. 2023. Are large language models really good logical reasoners? a comprehensive evaluation from deductive, inductive and abductive views. *arXiv preprint arXiv:2306.09841*.

Nikhil Yadav, Omkar Kudale, Aditi Rao, Srishti Gupta, and Ajitkumar Shitole. 2021. Twitter sentiment analysis using supervised machine learning. In *Intelligent data communication technologies and internet of things: Proceedings of ICICI 2020*, pages 631–642. Springer.