http://doi.org/10.35784/iapgos.6679

received: 09.10.2024 | revised: 28.10.2024 | accepted: 03.11.2024 | available online: 21.12.2024

EXTRACTING EMOTION-CAUSE PAIRS: A BILSTM-DRIVEN METHODOLOGY

Raga Madhuri Chandra, Giri Venkata Sai Tej Neelaiahgari, Satya Sumanth Vanapalli

Velagapudi Ramakrishna Siddhartha Engineering College, Department of Computer Science and Engineering, Vijayawada, India

Abstract. *Emotions are fundamental to human interactions, intricately influencing communication, behavior, and perception. Emotion-Cause Pair Extraction (ECPE) is a critical task in natural language processing that identifies clause pairs associating emotions with their corresponding triggers within textual documents. Unlike traditional Emotion Cause Extraction (ECE), which relies on pre-annotated emotion clauses, our study introduces a novel end-to-end model for ECPE. This innovative approach utilizes the extensive NTCIR-13 English Corpus to establish a robust baseline for ECPE in English, showcasing significant performance improvements over conventional multi-stage methods. Central to our model is the incorporation of Bidirectional Long Short-Term Memory (BiLSTM) networks, enhancing the ability to capture both local and global dependencies in textual sequences. By effectively combining contextual and positional embeddings, our model accurately predicts emotion-cause relationships, paving the way for a deeper understanding of emotional dynamics in conversational contexts and facilitating causal inference. Furthermore, our research highlights superior performance metrics, aligning its efficacy with state-of-the-art techniques in the field. This study advances emotion recognition in natural language processing, providing valuable insights for nuanced analyses of human emotions within textual data. Additionally, our findings enhance understanding of emotional intelligence in user interaction modeling and conversational AI applications. Through the public availability of our dataset and model, we aim to foster collaboration and further research in this vital area, ultimately improving the capacity for emotional understanding in applications ranging from sentiment analysis to interactive learning.*

Keywords: emotional intelligence, Emotion-Cause Pair Extraction (ECPE), Bidirectional Long Short-Term Memory (BiLSTM), emotional dynamics, Natural Language Processing (NLP), conversational analysis

WYODRĘBNIANIE PAR EMOCJA-PRZYCZYNA: METODOLOGIA OPARTA NA BiLSTM

Streszczenie. *Emocje mają fundamentalne znaczenie dla interakcji międzyludzkich, ściśle wpływając na komunikację, zachowanie i percepcję. Wyodrębnianie par emocja-przyczyna (ECPE) jest krytycznym zadaniem w przetwarzaniu języka naturalnego, które identyfikuje pary klauzul kojarzące emocje z odpowiadającymi im wyzwalaczami w dokumentach tekstowych. W przeciwieństwie do tradycyjnego wyodrębniania przyczyn emocji (ECE), które opiera się na wstępnie przypisanych klauzulach emocji, proponowane rozwiązanie wprowadza nowatorski kompleksowy model ECPE. To innowacyjne podejście wykorzystuje obszerny anglojęzyczny zbiór NTCIR-13 do ustanowienia solidnej podstawy dla ECPE w języku angielskim, wykazując znaczną poprawę wydajności w porównaniu z konwencjonalnymi metodami wieloetapowymi. Centralnym elementem modelu jest włączenie dwukierunkowych sieci pamięci długotrwałej (BiLSTM), co zwiększa zdolność do wychwytywania zarówno lokalnych, jak i globalnych zależności w sekwencjach tekstowych. Skutecznie łącząc osadzanie kontekstowe i pozycyjne, nasz model dokładnie przewiduje relacje emocji i przyczyn, torując drogę do głębszego zrozumienia dynamiki emocjonalnej w kontekstach konwersacyjnych i ułatwiając wnioskowanie przyczynowe. Co więcej, nasze badania podkreślają doskonałe wskaźniki wydajności, dostosowując ich skuteczność do najnowocześniejszych technik w tej dziedzinie. Badanie to rozwija rozpoznawanie emocji w przetwarzaniu języka naturalnego, dostarczając cennych spostrzeżeń dla zniuansowanych analiz ludzkich emocji w danych tekstowych. Ponadto nasze odkrycia zwiększają zrozumienie inteligencji emocjonalnej w modelowaniu interakcji użytkownika i konwersacyjnych aplikacjach AI. Poprzez publiczną dostępność naszego zbioru danych i modelu, dążymy do wspierania współpracy i dalszych badań w tym istotnym obszarze, ostatecznie poprawiając zdolność rozumienia emocji w zastosowaniach, od analizy nastrojów po interaktywne uczenie się.*

Słowa kluczowe: inteligencja emocjonalna, wyodrębnianie par emocja-przyczyna (ECPE), dwukierunkowa długa pamięć krótkotrwała (BiLSTM), dynamika emocjonalna, przetwarzanie języka naturalnego (NLP), analiza konwersacyjna

Introduction

The prediction of emotions from text [1, 22, 24] and the generation of emotionally-driven content [6, 13, 18, 21] have been widely studied, leading to significant advancements in understanding and manipulating emotional expression in writing. A key area of interest that emerged from this research is Emotion Cause Extraction (ECE), which focuses on identifying the underlying causes of emotions within a given text. Gui et al. (2016) were pioneers in this domain, introducing ECE as a method for breaking down sentences into clauses to detect which clause conveys the cause of a particular emotion. The challenge in ECE models, however, lies in their heavy reliance on emotion annotations for accurate predictions, making it difficult to test and apply them in various real-world scenarios where such annotations are absent.In 2019, Xia and Ding [25] took this further by introducing the Emotion-Cause Pair Extraction (ECPE) task, an extension of ECE that no longer depends on predefined emotion annotations. ECPE aims to identify all potential emotion-cause pairs within a document, broadening the application of sentiment analysis to more dynamic and real-time contexts. This advancement allows for the extraction of emotion causes from a wide array of sources, such as social media posts, product reviews, and other opinionated texts, without needing explicit emotion labels. By doing so, the ECPE task opens new possibilities for understanding sentiment in diverse domains. As depicted in Fig. 1, the annotated ground truth visually demonstrates how emotion-cause pairs are extracted, offering a clearer representation of the practical application of this method. This innovation marks a significant shift in sentiment research, enhancing the ability to analyze the roots of emotions in natural language across various contexts.

In the example provided, Clause 2 ("receiving the rejection email") serves as the cause clause, reflecting the source of emotions like disappointment and frustration, while Clause 4 ("her friend giving her concert tickets") functions as the cause clause for feelings of gratitude and joy expressed in Clause 3. Xia and Ding [25] proposed an approach to identifying potential cause and emotion clauses by initially separating these clauses into independent groups, where emotional clauses and their corresponding cause clauses were paired together, followed by filtering out invalid combinations. However, this method had limitations, particularly because it did not fully account for the interaction between emotion-expressing clauses and their related cause clauses. The failure to capture this interdependence could result in lower overall performance, especially when certain emotion clauses go unrecognized due to their undetected cause clauses, which are crucial for understanding the emotional context. In 2020, Ding et al. [8, 9] addressed these shortcomings by designing a more integrated framework that combines a robust encoder and classifier to enhance the detection of emotion-cause relationships. This unified approach not only improved performance but also emphasized the critical need to capture the interdependence between emotion and cause clauses in ECPE tasks, highlighting the importance of analyzing both aspects together to avoid missing essential context.

Clause 1: Emily received a rejection email from her dream job application.
Clause 2: She felt disappointment and frustration as she read the message.
Clause 3: Later that day, her friend surprised her with tickets to a concert.
Clause 4: Emily was filled with excitement and gratitude.

Fig. 1. A Sample Document

In this study, we introduce a complete framework for the Emotion-Cause Pair Extraction (ECPE) task, leveraging cooperative training to effectively simulate the link between emotions and their underlying causes. Our method simultaneously optimizes both elements, which leads to a more accurate emotion-cause pair extraction than existing multi-stage methods. The model outperformed traditional techniques, especially in tasks involving ground truth emotion annotations, when tested on a modified English-language dataset frequently used in sentiment research. By utilizing cutting-edge methods, our system sets new performance benchmarks by outperforming earlier Emotion-Cause Extraction models on important measures. We have made our dataset and model publicly accessible on GitHub to promote additional study and cooperation. This serves as a resource for creativity and validation among researchers.

1. Literature review

Wei Li et al. [16] introduced the task of emotion-cause pair extraction in conversations (ECPEC), aiming to identify emotional utterances and their causal contexts within dialogues. This work addresses challenges in conversational sentiment analysis (CSA) and emotion-cause pair extraction (ECPE) by employing causal inference techniques to establish links between emotional statements across conversational turns and neutral segments. The ECPEC task poses additional challenges due to the increased distance between emotional expressions and their causes in dialogues compared to traditional clause-level ECPE tasks. The authors developed the Conv ECPE dataset and proposed a two-step architecture tailored to this task. Experimental results on the Conv ECPE dataset demonstrated that this approach captures causal relationships in a way that enhances the analysis of conversational data, providing a benchmark for future work in conversational ECPE.

Fang Chen et al. [3] employed a recurrent neural network to synchronize emotions and causes within textual content, a method that forms the basis for this study's extension using BiLSTM for emotion-cause pair identification in real interactions. By examining emotion triggers in dynamic contexts, this work aims to improve applications in human-computer interaction (HCI) and consumer behaviour analysis. While Chen's approach effectively pairs emotions and causes, our BiLSTM-based method aims to refine this process further by capturing nuanced emotional dependencies in complex conversations, thus enhancing ECPE's scope and its applicability in diverse domains, including social media sentiment analysis and personalized marketing.

Chuang Fan et al. [10] proposed a transition-based approach for ECPE in emotion analysis, tackling limitations of traditional pipelined frameworks by employing a directed graph construction method. This model identifies emotions and causes simultaneously, optimizing subtasks jointly, thereby offering a deeper understanding of emotional dynamics within text. The performance improvements achieved by this approach demonstrate its advantage in capturing the interdependent nature of emotions and causes. This method's contributions underscore the significance of real-time emotion-driven applications across fields such as healthcare, customer service, and digital communication, where an accurate understanding of emotions is crucial for responsive system design.

Zifeng Cheng et al. [5] presented the Symmetric Local Search Network (SLSN) model, which aims to streamline ECPE by aligning cause clauses and emotions in tandem. Through two symmetric subnetworks, combining local pair searchers and clause representation learners, SLSN exhibits superior performance over earlier ECPE approaches, as evidenced by its results on the ECPE corpus. This dual-network approach manages complex linguistic structures and nuanced relationships between emotions and causes, establishing a benchmark in computational linguistics for its effectiveness in handling sophisticated language data in ECPE tasks.

Xinhong Chen et al. [4] advanced ECPE research by framing it as a sequence labelling problem, wherein their unified model assigned emotion type labels to clauses using a combination of CNN and BiLSTM networks. Their approach excelled in extracting multiple emotion-cause pairs within documents. Extending from this, our study applies BiLSTM to actual conversational datasets to investigate emotion-cause dependencies. This shift from document-based to dialogue-based ECPE represents a leap forward for NLP applications, facilitating improved insights in HCI and consumer behaviour research by enabling a richer understanding of emotional contexts within dialogues, benefiting applications in recommendation systems and sentiment analysis.

Dzmitry Bahdanau et al. [2] contributed a pivotal method in neural machine translation, introducing a soft-search mechanism in encoder-decoder models that identifies relevant portions of source phrases during translation. This adaptive approach enhances translation quality, showing strong results for English-to-French tasks. Although their method's primary application is in machine translation, it provides valuable insights for ECPE by demonstrating how models can capture contextual information without explicit segmentation, a principle that could inspire more advanced models in ECPE by refining contextual awareness.

Ashish Vaswani et al. [23] introduced the Transformer, a model that replaced recurrent and convolutional networks with attention mechanisms. This architecture significantly enhances parallelization capabilities, improving training efficiency and performance, as demonstrated on machine translation tasks. While the Transformer has advanced language modelling, its resource requirements pose challenges for ECPE applications where efficient processing of long sequences is essential. Nonetheless, the attention mechanism in Transformers highlights potential pathways for future ECPE research, where the ability to manage dependencies across lengthy conversational data can drive significant advancements.

J. Pennington et al. [20] introduced the GloVe (Global Vectors) model, which leverages global word co-occurrence statistics to generate word embeddings. By integrating matrix factorization and context window methods, GloVe offers robust performance on tasks like word analogy and named entity recognition. Although GloVe achieves efficient representation learning, its performance may be limited when capturing nuanced meanings from infrequent word co-occurrences, an area where alternative embeddings could offer improved accuracy. Nonetheless, GloVe's efficiency has informed word embedding practices across NLP, including tasks like ECPE where embedding quality affects model sensitivity to subtle emotional triggers.

K. Gao et al. [11] introduced a rule-based framework for emotion cause component detection within Chinese micro-blogs. This approach leverages a lexicon and Bayesian probability for identifying emotional triggers and causes, yielding valuable findings on emotion distribution across diverse linguistic contexts. While the rule-based nature of this model restricts its flexibility, the study highlights how rule-based techniques can support NLP tasks requiring linguistic precision, offering a foundational methodology for studies focusing on causal emotional analysis within culturally specific contexts.

Jacob Devlin et al. [7] presented BERT, a model that pre-trains deep bidirectional representations, capturing left and right context in a novel way. BERT surpasses prior models in multiple NLP tasks, achieving high scores on benchmarks like GLUE, MultiNLI, and SQuAD. Although BERT's extensive training requirements may present limitations in smaller-scale applications, its bidirectional attention mechanism offers a robust framework for advancing ECPE, allowing models to capture complex causal dependencies in conversational data. By integrating elements of BERT's bidirectionality, ECPE methods could evolve to handle more intricate conversational dynamics and sentiment analysis.

2. Proposed methodology

2.1. Task definition

Consider a document made up of an ordered sequence of clauses, denoted as $D = [x_1, x_2, ..., x_n]$. In ECPE tasks, pairs of emotion and cause clauses are represented as $P = \{..., (E_i, EC_j), ...\}$, where E_i signifies an emotion clause, and EC_j denotes the corresponding cause clause located within the document D. The primary objective of the Emotion-Cause Extraction task is to accurately identify and extract the cause clauses that are associated with each annotated emotion clause.

To illustrate the ECPE process, we present a visual summary in Fig. 2. This figure effectively demonstrates the methodology for locating and pairing emotion clauses within a document alongside their respective cause clauses. It highlights the intricate steps involved in this extraction process, particularly emphasizing how to achieve these pairings without relying on pre-established emotion annotations or labels. The visual representation serves as an invaluable aid, facilitating a better understanding of the ECPE task's techniques and methodologies. Through this visual assistance, readers can more easily grasp the systematic approach taken to identify and pair emotions with their corresponding causes, thereby enriching their comprehension of the underlying principles guiding the ECPE process.

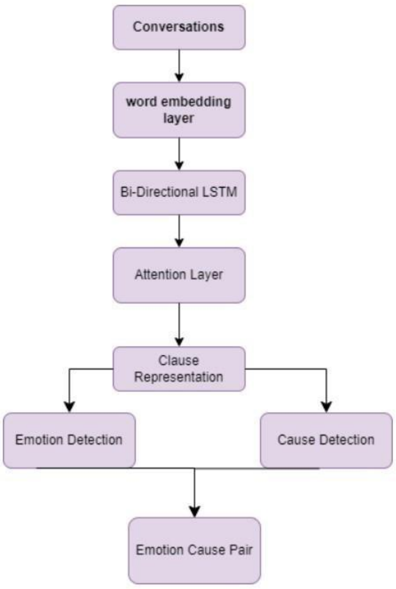

Fig. 2. Process Flow model

2.2. Proposed model

The proposed model for extracting emotion-cause pairs effectively identifies the complex connections between emotions and their causes in textual data. It processes entire documents to evaluate each clause pair for potential emotion-cause relationships. To enhance clause representations, the model is trained to detect emotions and identify causes through a BiLSTM network, benefiting from deep contextualized word representations as described by Peters et al. (2018) [19]. Contextual information from neighbouring clauses aids in predicting emotions and causes, with the Cause-Encoder utilizing these emotion predictions to refine its representations. By combining contextualized representations with positional embeddings, the model accurately identifies emotion-cause pairs, providing valuable insights into emotional dynamics and causal relationships in text.

2.3. Data collection

Our exploration into data collection begins with the Emotion-Cause Extraction (ECE) corpus, illustrated in Fig. 3, which offers insights into various emotion categories. Initially introduced at the renowned NTCIR-13 Workshop, this meticulously curated corpus is specifically designed to address the complexities of the ECE task. It includes a rich collection of 2,843 documents drawn from a diverse range of English novels, each annotated with great care. These annotations go beyond simple identification, exploring the intricate interplay of emotion-cause pairs woven within the text. Additionally, they specify the emotion category assigned to each clause, along with its identifying keyword, providing a detailed view of the emotional landscape depicted across the pages.

Fig. 3. Distribution of Emotion Categories

Throughout our training endeavours, our unwavering focus remains steadfastly fixated on unravelling the essence of emotion-cause pairs. This dedicated approach ensures a streamlined methodology, meticulously honed to exclude extraneous details and sharpen our focus on the primary objective at hand. Additionally, our exploration is complemented by Fig. 4, which illustrates the depiction of emotion clauses within the broader English corpus, further enriching our understanding of emotional dynamics and their representation in textual data.

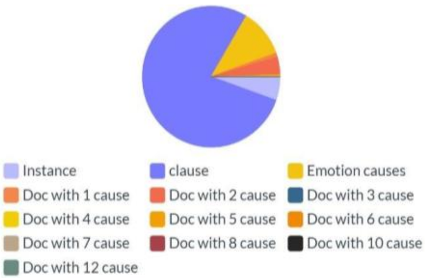

Fig. 4. Analysis of Emotion Causes

2.4. Data preprocessing

Data preprocessing plays an essential role in transforming raw textual data into a structured format suitable for analysis and modelling. It involves a series of interconnected tasks designed to improve data quality, consistency, and reliability, ensuring that the data can be effectively used in subsequent stages of analysis. The process starts with text cleaning, which focuses on removing noise and irrelevant information. This includes tasks such as eliminating special characters, punctuation, and HTML tags, as well as converting text into a standardized format by correcting spelling errors, expanding abbreviations, and handling non-standard encoding formats.

After cleaning, tokenization is applied to break the text into smaller, manageable units called tokens. These tokens, which can be words, phrases, or even sentences, serve as the fundamental components for natural language processing tasks. The tokenization process can vary depending on the goal, either at the word or sentence level, and is often guided by specific patterns or regular expressions to ensure that the tokens capture meaningful parts of the text. This organized and cleaned dataset then provides a solid foundation for further analysis, ensuring that models can efficiently process and interpret the data, ultimately improving the overall performance of the system.

The final step in the process is normalization, where the tokenized text is standardized to further improve consistency across the dataset. This involves converting all text to lowercase to avoid case sensitivity issues and removing stop words to reduce the size of the vocabulary, making the data more manageable. Techniques like stemming and lemmatization are applied to reduce words to their root forms, preserving their semantic meaning and reducing dimensionality, which is critical for effective text analysis. By performing these preprocessing steps in sequence, the raw textual data is transformed into a structured and standardized format, facilitating more accurate and reliable results in natural language processing tasks.

3. Approach

We propose a specialized model, E2E-PExtE, designed to comprehensively extract emotion-cause pairs from entire documents. The model evaluates the likelihood that pairs of clauses (ci, cj) represent potential emotion-cause relationships, thereby uncovering the intricate connections between emotional states and their underlying causes within textual data. E2E-PExtE employs a hierarchical architecture that integrates various neural network components, which enhance its ability to process and analyze complex textual information.

At the heart of its architecture are BiLSTM networks that operate at the word level, enabling the model to capture semantic nuances by modelling bidirectional sequential dependencies. This bidirectional nature allows the model to account for both preceding and succeeding words, assimilating contextual cues from surrounding text. By doing so, the model generates refined word-level representations, where each word's meaning and context are encapsulated, thereby enriching the overall understanding of the document. This hierarchical structure empowers the model to not only identify emotion-cause pairs effectively but also improve its ability to interpret the semantic relationships within the text.

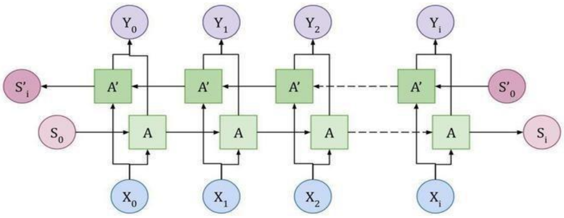

Fig. 5. BiLSTM Architecture

The architecture of the Bidirectional Long Short-Term Memory (BiLSTM) layer plays a key role in how the model processes textual sequences, as illustrated in Fig. 5. In this architecture, the input token (X_i) and output token (Y_i) are processed through LSTM nodes (A and A'), with two LSTM networks working in tandem to process the sequence in both forward and backward directions. This bidirectional structure is essential, as it captures both preceding and succeeding contexts, thereby enriching word representations with information from both the past and future.

The hidden states generated from each direction are concatenated, resulting in context-aware word representations. These representations encapsulate not only the meaning of individual words but also their contextual relevance within the larger text, which is then forwarded for further analysis in subsequent neural network layers. Such a mechanism significantly enhances the model's understanding of the intricate dynamics present in natural language.

The BiLSTM layer described above is central to the emotion-cause pair extraction process. By leveraging the power of bidirectional processing, this layer enhances the model's ability to learn from both past and future contexts in a conversation. As a result, it improves the accuracy of detecting relationships between emotions and their causes, allowing the model to generate more precise predictions regarding emotional states and their triggers. This capability is particularly vital in understanding nuanced emotional expressions in dialogue, where the sentiment may depend heavily on context. Below, Algorithm 1 outlines the process of emotion detection using this BiLSTM architecture, demonstrating its critical role in the overall framework we have developed for emotion-cause pair extraction.

Algorithm 1: Emotion Detection from Conversations

Input: A conversation dataset D, where each conversation is a sequence of sentences or clauses D= $\{S_1,S_2.....,S_n\}$; Pre-trained word embeddings E (optional).

Output: Detected emotion labels $\{e_1,e_2......,e_n\}$ for each sentence, where $e_i \in$ {happy, sad, angry, surprised, etc.}.

1. Tokenize the conversation into sentences or clauses, clean text, and convert each tokenized sentence S_i into numerical embeddings E.
2. Initialize a BiLSTM network, input the embedded sequence of S_i and capture the forward and backward dependencies using the BiLSTM model.
3. Apply a dense layer with softmax activation to output the probability distribution over emotions for each S_i.
4. Train the model with categorical cross-entropy and output the most likely emotion label for each sentence S_i.

At the clause level, the BiLSTM layer encodes sequential information, capturing both local and global dependencies. The Emotion-Encoder and Cause-Encoder, both BiLSTM-based, collaborate by informing each other's predictions to better identify causal links. For emotion-cause pair extraction, the model combines contextual and positional embeddings, using BiLSTM networks to predict emotion-cause relationships. Neural layers further refine these predictions, leveraging emotional content and positional cues for accurate extraction, enhancing the understanding of emotional dynamics in text.

This process is summarized in Algorithm 2, which outlines the method used for emotion-cause pair extraction following the detection of emotions in sentences or clauses. In this algorithm, a list of sentences with detected emotions is processed alongside their context sentences to identify potential causes of each emotion. The algorithm begins by extracting context from adjacent sentences, forming pairs that facilitate the interaction modelling between the target sentence and its context. Utilizing a BiLSTM network, the model assesses these interactions and employs a binary classification layer to discern which context sentences serve as causes for the identified emotions. This structured approach not only streamlines the extraction process but also enhances the model's capability to capture the intricate relationships between emotions and their causes within textual data.

Algorithm 2: Cause Detection for Emotion-Cause Pair Extraction

Input: A list of sentences with detected emotions $\{(S_1,e_1),(S_2\ ,e_2),\ldots,(S_n,e_n)\}$ from Phase 1;Context sentences from conversation dataset D.

Output: Causes $\{c_1,c_2,\ldots,c_n\}$, where each c_i represents the cause of emotion e_i in S_i .

1. For each S_i with emotion e_i , extract context $\{S_{i-1},S_{i+1}\}$ and form pairs $\{(S_i,S_{i-1}),(S_i,S_{i+1})\}$.
2. Initialize a BiLSTM network to model the interaction between S_i and its context.
3. Use a dense layer with binary classification (cause or not-cause), training with binary cross-entropy loss.
4. Output the context sentence(s) identified as the cause c_i of emotion e_i.

The mathematical representation for fully connected layers is as follows:

$$y = \sigma(Wx + b)y = \sigma(Wx + b) \quad (1)$$

where:
x is the input vector, W is the weight matrix, b is the bias vector, σ is the activation function, such as ReLU or sigmoid, yy is the output vector.

In our model, fully connected networks are essential for extracting meaningful insights from complex text by learning intricate relationships between emotions and causes. These networks use dense connectivity to adjust weights during training, identifying significant features and patterns for accurate emotion-cause pair predictions. They enable the model to decode emotional dynamics and provide structured representations for in-depth analysis. The E2E-PExtE model optimizes its settings by integrating weighted losses from various tasks, addressing class imbalances, and adjusting loss weights, resulting in improved performance and more precise emotion-cause pair extraction.

The loss function for training the model can be defined as a weighted sum of the losses on the primary and auxiliary tasks:

$$Ltotal = \lambda c \cdot Lc + \lambda e \cdot Le + \lambda p \cdot Lp \quad (2)$$

where:
Ltotal is the total loss, Lc, Le, and Lp are the losses for cause prediction, emotion prediction, and pair prediction, respectively, c, λe, and λp are the weights for the respective losses.

Influenced by recent research, the E2E-PExtE model incorporates an innovative interactive task learning strategy designed to enhance its overall performance. A key feature of this approach is the functionality of the Cause-Encoder, which actively utilizes predictions generated by the Emotion-Encoder during the encoding phase of the model. This integration allows the model to infuse emotional insights directly into the encoding process of cause-clauses, employing BiLSTM networks to analyze and capture the intricate contextual nuances that exist within causal relationships. By enabling the Cause-Encoder to leverage the predictions made by the Emotion-Encoder, the model is adept at recognizing how emotions can influence or give rise to specific causes. This interactive learning mechanism fosters a deeper understanding of the connections between emotional expressions and their corresponding causes, significantly enhancing the model's ability to discern causality within textual data. As a result, the E2E-PExtE model achieves more precise pair extraction, allowing for a more accurate representation of emotional dynamics in text. This advancement highlights the model's capability to reflect the complexity of human emotional expression, ultimately leading to improved outcomes in emotion-cause pair extraction tasks.

4. Experiments and result

We conducted a comprehensive comparative analysis of our model against established benchmarks, including notable prior works such as the ECPE 2-stage model by Xia et al. [25], as well as the ECPE-2D(BERT) and ECPE-MLL(ISML-6) models developed by Ding et al. [8, 9]. This evaluation aimed to assess the performance of our model using our custom dataset, providing a thorough understanding of its efficacy in extracting emotion-cause pairs.

Training our model was executed using the Adam optimizer [20, 26] over 15 epochs, utilizing 200-dimensional GloVe word embeddings for rich word representation. For optimal performance, model weights and biases were initialized from a uniform distribution U(-0.10, 0.10), with a carefully selected learning rate of 0.005 to facilitate convergence. To combat overfitting and enhance generalization, we implemented regularization techniques, including L2 weight decay set at 1e-5 specifically for the softmax parameters, and a dropout rate of 0.8 for the word embeddings. In configuring the model, we assigned hyperparameters $\lambda c : \lambda e : \lambda p$ as $1:1:2.5$ to underscore the critical importance of accurately detecting pairs in the emotion-cause extraction process. Additionally, we employed randomly initialized embeddings and optimized positional embeddings through a clipping distance of 10, ensuring that the model could effectively learn spatial relationships within the data.

During the training process, we utilized the Adam optimizer [20, 26], which dynamically adjusts learning rates by leveraging exponentially decaying averages of past gradients. This adaptive optimization strategy not only accelerates convergence but also promotes stability compared to traditional methods. The key hyperparameters of the Adam optimizer – namely the learning rate of 0.005, along with beta1, beta2, and epsilon –are crucial in finely tuning the optimization process and improving model performance.

To further understand the impact of varying loss weights on the performance of our model, we conducted an analysis focusing on precision, recall, and F1 score under different configurations. The results of this analysis are visualized in Fig. 6 (Precision vs. Loss Weight), Fig. 7 (Recall vs. Loss Weight), and Fig. 8 (F1 Score vs. Loss Weight). These figures offer valuable insights into how adjusting loss weights can influence the balance between precision, recall, and F1 score, thereby providing a deeper understanding of model behaviour and performance dynamics under various training conditions.

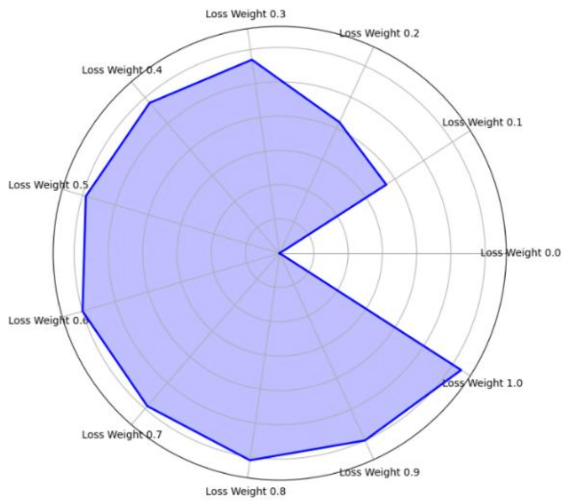

Fig. 6. Precision vs Loss Weight

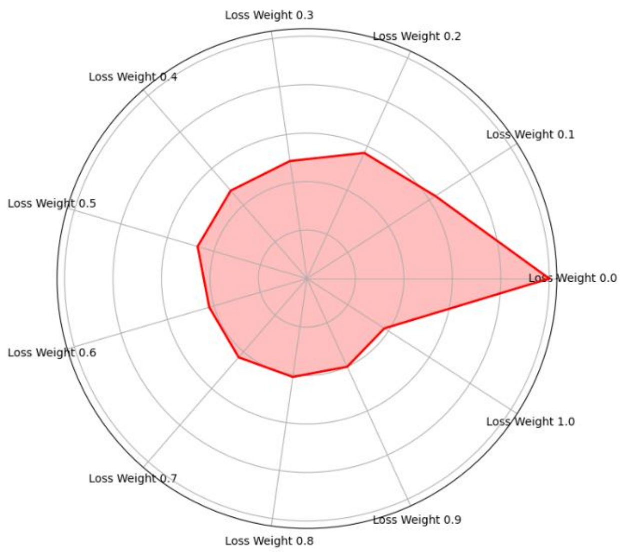

Fig. 7. Recall vs Loss Weight

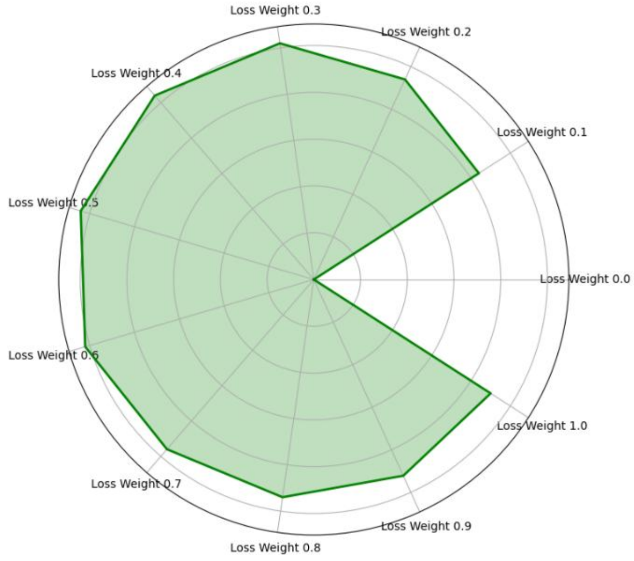

Fig. 8. F1-Score vs Loss Weight

Xia et al. [25] define the metrics as follows:

$$Precision \leftarrow \frac{True\ Positives}{True\ Positives+False\ Positives} \quad (2)$$

$$Recall \leftarrow \frac{True\ Positives}{True\ Positives+False\ Negatives} \quad (3)$$

$$F1\ score \leftarrow 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

We evaluated the model's performance using standard metrics such as F1-score, precision, and recall. Our end-to-end approach, which effectively utilizes the interdependence between emotion and cause clauses, demonstrated superior performance over the ECPE 2-stage model, achieving a significant improvement of 6.5% in emotion-cause pair extraction. This improvement is notable when compared to prior work in the field, such as the co-attention neural network model proposed by Li et al. (2018) [17], which emphasizes emotional context awareness in emotion-cause analysis. Their approach highlights the importance of contextual information, which aligns with our findings that interdependencies significantly enhance extraction accuracy.

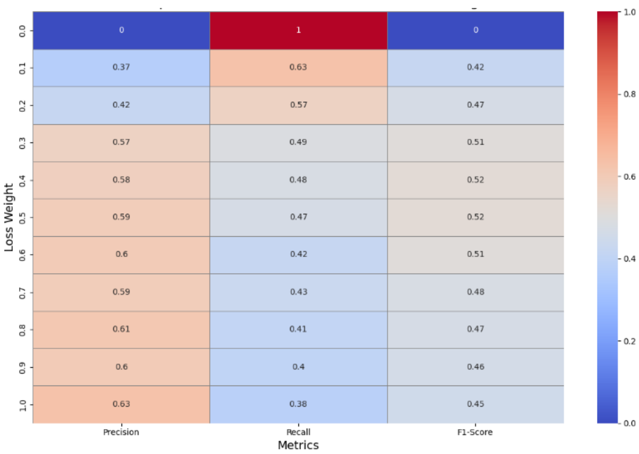

Fig. 9. Precision, Recall, F1 Score vs Loss Weight as a heatmap

Fig. 9 shows the Precision, Recall, and F1 Score vs. Loss Weight as a heatmap, illustrating the model's performance across different loss weights. Moreover, E2E-PExtE demonstrated performance comparable to advanced models like ECPE-MLL(ISML-6) and ECPE-2D(BERT) [8]. It also excelled in the Emotion-Cause Extraction (ECE) task, with its variant, E2E-CExt, outperforming the state-of-the-art RHNN model (Fan et al., 2019) [10], underscoring its generalizability and the importance of enhancing emotion predictions for better ECPE performance.

Fig. 10. User Interface for Emotion-Cause Pair Extraction – Example 1

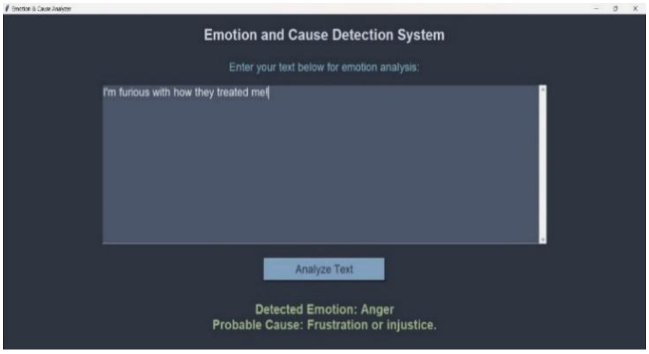

Fig. 11. User Interface for Emotion-Cause Pair Extraction – Example 2

To visually represent the performance and functionality of our E2E-PExtE model, Fig. 10 and Fig. 11 illustrate the User Interface for Emotion-Cause Pair Extraction, highlighting the model's ability to analyze and extract emotion-cause pairs from textual data in a user-friendly manner. In Fig. 10, users input a sentence such as "This news fills me with joy" and initiate the inference process. The interface identifies the emotion-cause pair, displaying happiness as the emotion and positive experience or good news as the cause. In Fig. 11, the input sentence "I'm furious with how they treated me!" prompts the interface to extract anger as the emotion and frustration or injustice as the cause. These examples demonstrate how the model processes complex textual data to uncover meaningful connections between emotions and their underlying causes, underscoring the effectiveness of our approach in real-world scenarios. The intuitive design of the interface enables seamless user engagement, facilitating the exploration of emotional dynamics in various contexts. This practical application illustrates the model's robustness and versatility, effectively decoding emotional expressions and their motivations, providing insights valuable in fields such as sentiment analysis, customer feedback evaluation, and mental health monitoring.

5. Conclusion and future work

This study advances emotion analysis and cause prediction using deep learning for real-time applications. Our approach accurately identifies causes behind emotional states through refined emotion-cause pair extraction techniques, validated on labeled datasets. By optimizing model architecture, hyperparameters, and dataset integration, we enhance generalization for emotional trigger detection. Future work involves integrating our system into real-world applications like social media analytics and customer feedback systems to provide insights that improve user experiences. This project highlights deep learning's potential in understanding human behavior and aims to foster innovative applications in emotion analysis.

References

[1] Alswaidan N., Menai M. E. B.: A survey of state-of-the-art approaches for emotion recognition in text. Knowledge and Information Systems 62(8), 2020, 2937–2987 [https://doi.org/10.1007/s10115-020-01449-0].
[2] Bahdanau D., Cho K., Bengio Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473. 2014.
[3] Chen F., Shi Z., Yang Z., Huang Y.: Recurrent synchronization network for emotion-cause pair extraction. Knowledge-Based Systems 238, 2022, 107965.
[4] Chen X., Li Q., Wang J.: A unified sequence labeling model for emotion cause pair extraction. 28th International Conference on Computational Linguistics, 2020, 208–218.
[5] Cheng Z., Jiang Z., Yin Y., Yu H., Gu Q.: A symmetric local search network for emotion-cause pair extraction. 28th International Conference on Computational Linguistics, 2020, 139–149.
[6] Colombo P., Witon W., Modi A., Kennedy J., Kapadia M.: Affect-driven dialog generation. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minnesota, Minneapolis 2019, 3734–3743 [https://doi.org/10.18653/v1/N19-1374].
[7] Devlin J.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
[8] Ding Z., Xia R., Yu J.: ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. 58th Annual Meeting of the Association for Computational Linguistics, 2020, 3161–3170 [https://doi.org/10.18653/v1/2020.acl-main.288].
[9] Ding Z., Xia R., Yu J.: End-to-end emotion-cause pair extraction based on sliding window multi-label learning. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, 3574–3583.
[10] Fan C., Yuan C., Du J., Gui L., Yang M., Xu R.: Transition-based directed graph construction for emotion-cause pair extraction. 58th Annual Meeting of the Association for Computational Linguistics, 2020, 3707–3717.
[11] Gao K., Xu H., Wang J.: A rule-based approach to emotion cause detection for chinese micro-blogs. Expert Syst. Appl. 42(9), 2015, 4517–4528.
[12] Gao Q., Hu J., Xu R., Gui L., He Y., Wong K.-F., Lu Q.: Overview of NTCIR-13 ECA Task. 13th NTCIR Conference on Evaluation of Information Access Technologies. Japan, Tokyo 2017, 361–366.
[13] Ghosh S., Chollet M., Laksana E., Morency L. P., Scherer S.: Affect-LM: A neural language model for customizable affective text generation. 55th Annual Meeting of the Association for Computational Linguistics. Canada, Vancouver 2017, 634–642 [https://doi.org/10.18653/v1/P17-1059].
[14] Gui Lin, Xu R., Wu D., Lu Q., Zhou Y.: Event-driven emotion cause extraction with corpus construction. Wong K.-F. et al. (eds): Social Media Content Analysis: Natural Language Processing and Beyond. World Scientific, 2018, 145–160.
[15] Kingma D. P., Ba J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
[16] Li W. et al.: ECPEC: Emotion-Cause Pair Extraction in Conversations. IEEE Transactions on Affective Computing 14(03), 2023, 1754–1765.
[17] Li X., Song K., Feng S., Wang D., Zhang Y.: A co-attention neural network model for emotion cause analysis with emotional context awareness. Conference on Empirical Methods in Natural Language Processing, 2018, 4752–4757.
[18] Modi A., Kapadia M., Fidaleo D. A., Kennedy J. R., Witon W., Colombo P.: Affect-driven dialog generation. U.S. Patent 10,818,312 B2, October 27, 2020.
[19] Neumann M. P. M., Iyyer M., Gardner M., Clark C., Lee K., Zettlemoyer L.: Deep contextualized word representations. arXiv preprint arXiv:1802.05365, 2018.
[20] Pennington J., Socher R., Manning C. D.: Glove: Global vectors for word representation. Conference on Empirical Methods in Natural Language processing – EMNLP, 2014, 1532–1543.
[21] Singh I., Barkati A., Goswamy T., Modi A.: Adapting a language model for controlled affective text generation. arXiv preprint arXiv:2011.04000 (2020).
[22] Strapparava C., Mihalcea R.: Learning to identify emotions in text. ACM symposium on Applied computing, 2008, 1556–1560.
[23] Vaswani A. et al.: Attention is all you need. Advances in neural information processing systems. 30th Advances in Neural Information Processing Systems – NIPS, 2017, 5998–6008.
[24] Witon W., Colombo P., Modi A., Kapadia M.: Disney at IEST 2018: Predicting emotions using an ensemble. 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Belgium, Brussels 2018, 248–253 [https://doi.org/10.18653/v1/P17].
[25] Xia R., Ding Z.: Emotion-cause pair extraction: A new task to emotion analysis in texts. arXiv preprint arXiv:1906.01267 (2019).
[26] Zhang Z.: Improved adam optimizer for deep neural networks. IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), 2018.

Asst. Prof. Raga Madhuri Chandra
e-mail: madhuri.chandra1209@gmail.com

Raga Madhuri Chandra is currently serving as an assistant professor in the Department of Computer Science and Engineering at Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India. She completed her M.Tech. in Computer Science and Engineering in 2014 from Jawaharlal Nehru Technological University, Kakinada, and her B.Tech. in Information Technology in 2012 from the same university. With over 8 years of teaching experience, her research interests encompass Data Mining, Machine Learning, Deep Learning, and Cloud Computing. She has contributed significantly to her field through multiple publications in national and international journals and holds 7 patents related to her research work.

https://orcid.org/0000-0002-3687-0783

Eng. Giri Venkata Sai Tej Neelaiahgari
e-mail: ngvst2004@gmail.com

Giri Venkata Sai Tej Neelaiahgari is a fourth-year B. Tech. student specializing in Computer Science and Engineering at V.R.Siddhartha Engineering College, Vijayawada, India. He is passionate about Machine Learning, the Internet of Things, and Web Programming.

https://orcid.org/0009-0009-5785-0413

Eng. Satya Sumanth Vanapalli
e-mail: sumanthvanapalli074@gmail.com

Satya Sumanth Vanapalli is a fourth-year B.Tech. student specializing in Computer Science and Engineering at V.R. Siddhartha Engineering College, Vijayawada, India. His interests lie in the fields of Internet of Things, Machine Learning and Web Programming.

https://orcid.org/0009-0006-5993-3456