End-to-End Emotion Semantic Parsing

Xiaotong Jiang, Zhongqing Wang*, and Guodong Zhou
[1]Natural Language Processing Lab, Soochow University, Suzhou, China
xtjiang21@stu.suda.edu.cn,{wangzq,gdzhou}@suda.edu.cn

Abstract

Emotion detection is the task of automatically associating one or more emotions with a text. The emotions are experienced, targeted, and caused by different semantic constituents. Therefore, it is necessary to incorporate these semantic constituents into the process of emotion detection. In this study, we propose a new task called *emotion semantic parsing* which aims to parse the emotion and semantic constituents into an abstract semantic tree structure. In particular, we design an end-to-end generation model to capture the relations between emotion and all the semantic constituents, and to generate them jointly. Furthermore, we employ a task decomposition strategy to capture the semantic relation among these constituents in a more cognitive and structural way. Experimental results demonstrate the importance of the proposed task, and indicate the proposed model gives superior performance compared to other models.

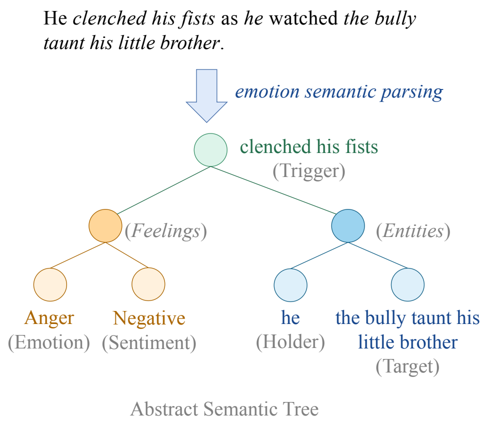

Figure 1: Example of abstract semantic tree.

1 Introduction

The goal of emotion detection is to automatically detect or categorize the emotional states of human according to some inputs. Nowadays, emotion detection can be found in a broad range of applications, including but not limited to emotional support (Tu et al., 2022; Pavarini et al., 2023), human-computer interaction (Chowdary et al., 2021; Alrowais et al., 2023) and healthcare surveillance (Dhuheir et al., 2021; Fang et al., 2023). Henceforth, emotion detection has attracted increasing attention from both research community and industry in recent years (Hu et al., 2021; Zanwar et al., 2023).

In literature, most studies focus on detecting emotion from the given text. For example, feature-based works (Balahur et al., 2011) and deep learning approaches, including RNNs (Majumder et al., 2019), Transformers (Yu et al., 2018), and pre-trained language models (Zheng et al., 2022; Zanwar et al., 2023) are used in emotion detection. Nevertheless, over the past few years, numerous studies (Xia and Ding, 2019; Kim and Klinger, 2018; Campagnano et al., 2022) have indicated that the emotions may be experienced, targeted, and/or caused by different semantic constituents, and can be linked to form abstract semantic structures.

However, most existing works only take "cause" into consideration, neglecting the intricate relationships between emotion constituents, such as the strong relationship between trigger and emotion, as well as the similarity of sentiment and emotion labels. This limitation makes acquiring a nuanced understanding of the interplay between emotion and semantic constituents challenging, thereby hindering the model's capability to thoroughly comprehend emotional expression. By addressing these complexities, we believe it can serve as a foundational component for improving results in downstream real-world tasks by providing a more comprehensive understanding of emotional dynamics and their influence.

To tackle the above problem, we propose a new

* Corresponding author

Findings of the Association for Computational Linguistics ACL 2024, pages 37–47
August 11-16, 2024

task called *emotion semantic parsing* which aims to parse the emotion and semantic constituents into an abstract semantic tree structure. In this work, we consider five emotion-related semantic constituents into this task, including *emotion*, *sentiment*, *trigger*, *holder*, and *target*. As shown in Figure 1, emotion and sentiment refer to the specific emotion and sentiment labels of the sentence; trigger refers to a trigger word or expression that describes an emotion; and the constituents of holder and target correspond to the person or entities towards whom or from whom the emotion is directed, respectively.

Moreover, we use the form of the *abstract semantic tree structure* to model the relationship between these constituents, we use the Trigger node to represent the whole abstract semantic tree, and we use the virtual Feelings and Entities nodes to compass other constituent nodes. The tree structure can be considered as a semantic representation in order to better represent the structure of the emotion-related constituents. As shown in Figure 1, the tree structure models a sentence using a rooted directed acyclic graph, highlighting its main elements (e.g. emotion, triggers) and semantic relations. It can thus potentially reveal a more comprehensive and complete semantic structure for extracting these constituents. Furthermore, since the emotion and semantic constituents are correlated in the form of the abstract semantic tree structure, we design an end-to-end generative model to predict them jointly. Specifically, our proposed model begins by decomposing the original task into several sub-tasks, as shown in Figure 2. Subsequently, the proposed model utilizes template-based prompt learning to extract each semantic constituent and semantic constituents pair separately for each sub-tasks. Finally, the proposed model leverages the acquired knowledge from sub-tasks to generate all semantic constituents in the main task.

We conducted extensive experiments on the unified benchmark. Experimental results demonstrate the importance of the proposed task, and indicate the proposed model gives superior performance compared to other models. Further analysis shows the effect of the proposed abstract semantic tree structure and decomposition strategies.

2 Related Works

In the domain of emotion analysis, previous studies have mainly focused on text-based emotion detection. Early feature-based work often relied on traditional machine learning algorithms and manually crafted word-level emotion lexicons (Balahur et al., 2011). As deep neural networks progressed, emotion detection methods began utilizing various networks, such as RNNs (Majumder et al., 2019) and Transformers (Vaswani et al., 2017; Yu et al., 2018), to learn the emotion-related representation for detection. Recently, pre-trained language models (Devlin et al., 2019; Raffel et al., 2020) have dominated this area with its flexibility and rich internal knowledge (Alvarez-Gonzalez et al., 2021; Zheng et al., 2022).

The emotions are not expressed individually, and can be experienced, targeted, and caused by different *semantic constituents* (Xia and Ding, 2019; Xiao et al., 2023; Campagnano et al., 2022). Most recent studies focus on emotion cause extraction task (Xia and Ding, 2019; Wang et al., 2023; Gu et al., 2023), with the goal of extracting the potential pairs of emotions and corresponding causes in the document and conversation scenario. Many end-to-end approaches have been proposed. These works can be broadly categorized into two main groups: some focus on multi-task learning (Ding et al., 2020a,b; Chen et al., 2022), while others transform the original task into a unified sequence labeling problem (Yuan et al., 2020; Chen et al., 2020). Recently, Xiao et al. (2023) developed a rule-based system to extract emotion causes on social media from constituency parsing trees. Wang et al. (2023) proposed emotion cause triplet extraction to extract emotion-cause-category triplets for fine-grained emotion cause analysis.

Different from prior studies, we introduce a novel task, termed *emotion semantic parsing*. This task seeks to comprehensively examine the impact of various semantic constituents on emotional expressions. To achieve this, we develop an end-to-end generative model specifically designed to capture the intricate relationships between emotion and a wide range of semantic constituents.

3 Preliminaries

As shown in Figure 2, we introduce a novel task termed *emotion semantic parsing*. This task focuses on the generation of an abstract semantic tree structure that encapsulates both emotion and semantic constituents. To achieve this, we develop an end-to-end generative model specifically tailored for the generation of the abstract semantic tree structure. Our proposed model adopts a decomposition-based strategy, breaking down the complex task into several more manageable sub-tasks. Specifically, our proposed model begins by decomposing the original task into several sub-tasks. Finally, the proposed model leverages the acquired knowledge from sub-tasks to generate all semantic constituents in the main task.

In this section, we introduce two important issues of this task: abstract semantic tree construction and the abstract semantic tree parsing model. The decomposition-based strategy for abstract semantic tree parsing will be discussed in the next section.

3.1 Abstract Semantic Tree Construction

Inspired by the widespread impact of tree structures in parsing tasks (Li et al., 2023; Groschwitz et al., 2023), we represent the relationships among multiple constituents in a tree structure. This approach allows us to capture the complex dynamics inherent in emotional expression. We first introduce the constituents in the abstract semantic tree, then we describe the construction process of the tree.

Definitions of Semantic Constituents

There are five types of semantic constituents, each serving as a node in the abstract semantic tree. Their definitions are provided below:

- *Emotion* is the specific emotion label represented from the original text. In this study, we use Plutchik (2001)'s taxonomy of emotion.
- *Sentiment* is the specific sentiment label represented from the original text, including "positive", "negative", and "neutral".
- *Trigger* refers to a trigger word or expression that describes an emotion, even if it is implicit.
- *Holder* refers to a person or entity that feels or experiences the emotion identified by the trigger.
- *Target* refers to a person or entity towards whom/which the emotion identified by the trigger is directed.

Process of Tree Construction

As shown in Figure 1, the abstract semantic tree represents a sentence through a rooted directed acyclic graph utilizing the aforementioned semantic constituents. The construction process is outlined below:

- We take the "Trigger" node as the root node, since it is the most important constituent in the emotion parsing.
- The virtual nodes "Feelings" and "Entities" are linked to Trigger node.
- The "Sentiment" and "Emotion" are linked to "Feelings" as terminal nodes, while the "Holder" and "Target" are linked to "Entities" node.

As Figure 1 shows, the trigger "clenched his fists" implies the author's emotion. The holder and target are "he" and "the bully taunt his little brother", respectively. By leveraging these emotion-related semantic constituents, the model can accurately predict the "anger" emotion and "negative" polarity of the sentence.

3.2 Basic Semantic Tree Parsing

After we construct the abstract semantic tree, we then employ a generative model to parse the emotion and all the semantic constituents into a linearized tree structure. As shown in Figure 2, given an input text, the output can be represented as a linearized semantic tree.

We adopt the pre-trained language model T5 (Chung et al., 2022) as our transformer-based encoder-decoder architecture. Therefore, by formulating parsing tasks as a text generation problem, we can tackle them in a unified sequence-to-sequence framework without a task-specific model design.

The main challenge of emotion semantic parsing is the complex structure inherently in the emotional expression. Specifically, the complex structure will bring out two challenges: First, the large number of semantic constituents raises the complexity to the parsing task. Consequently, the precise extraction of every constituent becomes challenging yet indispensable. Second, effectively modeling these internal relationships also presents a significant challenge in the task.

4 Decomposition-based Semantic Tree Parsing Model

In this study, we propose a decomposed-based generative model to tackle the above challenges. The motivation is to alleviate the complexity of the original task by decomposing it into several sub-tasks. As shown in Figure 2, our proposed model can be

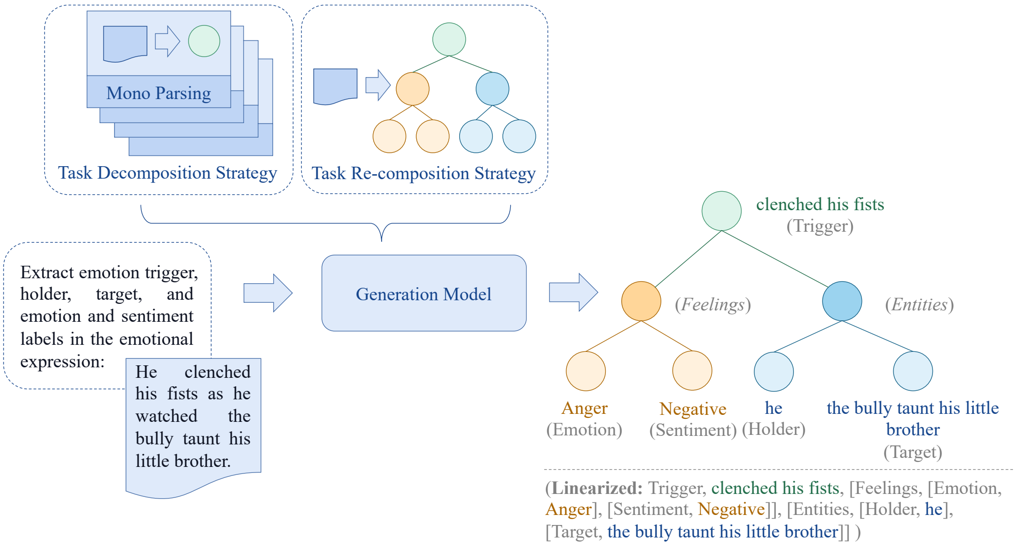

Figure 2: Overview of the proposed model.

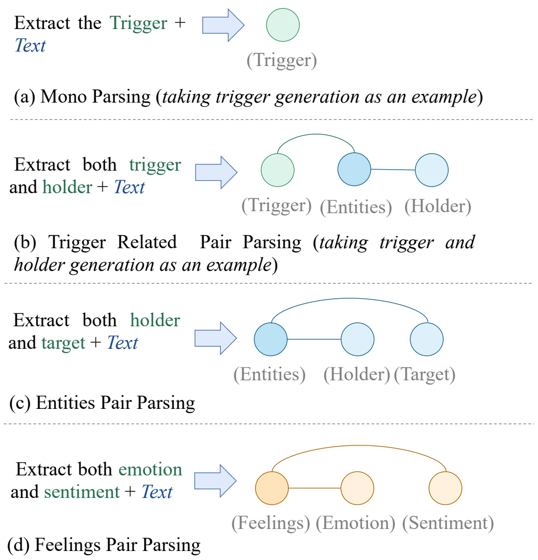

Figure 3: Examples of task decomposition strategies. Due to the space limit, prompt texts in this figure are shortened. Original prompt texts are presented in Section 4.1.

separated into two stages: decomposition and re-composition. Our model firstly breaks down the original abstract semantic parsing task into several sub-tasks, and then designs individual instruction-based templates for each task. All sub-tasks are fine-tuned by a generative backbone at the same time. Based on the model parameters from the last re-composition stage, our model further finetunes its parameters again with the original task, which involves jointly parsing all semantic constituents.

4.1 Task Decomposition Strategy

To implement a generalized framework that is applicable to all tasks, we design templates for each individual task. The input template for each task comprises the task-specific instruction, and the original text to be parsed. The output template for each task is the sub-tree containing the constituents to be parsed. The following templates use the example in Figure 2 as the *Original Text*.

Mono Parsing includes trigger, emotion, sentiment, holder, and target parsing, individually. As shown in Figure 3(a), these sub-tasks aim to capture each specific constituent:

> //taking emotion parsing task as an example
>
> `INPUT` Detect the emotion label in the emotional expression: + *Original Text*
>
> `Output` [Emotion (Anger)]

Trigger-related Pair Parsing includes trigger-emotion, trigger-sentiment, trigger-holder, trigger-target, trigger-feelings, and trigger-entities. As shown in Figure 3(b), these sub-tasks aim to capture the relationship between the trigger and each specific constituent:

> //taking trigger-holder parsing task as an example
>
> `INPUT` Extract both the trigger and holder in the emotional expression: + *Original Text*
>
> `Output` [Trigger (clench his fists), [Entities, [Holder (he)]]]

Entities Pair Parsing aims to capture the relationship between holder and target, and can gain insights into the interpersonal dynamics and relational aspects that influence emotional experiences. The example is shown in Figure 3(c), with the template of prompt provided below:

`INPUT` Extract both holder and target in the emotional expression: + *Original Text*

`Output` [Entities, [Holder (he), Target (the bully taunt his little brother)]]

Feelings Pair Parsing aims to capture the relationship between sentiment and emotion. The example is shown in Figure 3(d), with the template of prompt provided below:

`INPUT` Extract both the emotion and sentiment labels in the emotional expression: + *Original Text*

`Output` [Feelings, [Emotion (Anger), Sentiment (Negative)]]

4.2 Task Re-composition Strategy

After gaining knowledge from sub-tasks, the model has learned the pattern of the template and the relationships from different emotion-related semantic constituents. Thus it can boost the final end-to-end emotion semantic parsing task, which is to predict the emotion label and outputs all emotion-related semantic constituents in the form of an abstract semantic tree. The example is shown in Figure 2, and the template of prompt is the below:

`INPUT` Extract emotion trigger, holder, target, and emotion and sentiment labels in the emotional expression: + *Original Text*

`Output` [Trigger (clenched his fists), [Feelings, [Emotion (Anger), Sentiment (Negative)]], [Entities, [Holder (he), Target (the bully taunt his little brother)]]]

4.3 Objective Functions and Training

In this subsection, we show the objective functions and training process of the proposed model.

The goal is to maximize the output linearized tree X_T probability given the sentence X. Therefore, we optimize the negative log-likelihood loss function:

	Amount
#Sample	7,234
Avg. Sent Len.	26.3
#Emotion Labels	9
#Sentiment Label	3
#Trigger	7,111
#Holder	5,633
#Target	6,127

Table 1: Distribution of dataset.

$$\mathcal{L} = -\frac{1}{|\tau|} \sum_{(X,X_T)\in\tau} \log p(X_T|X;\theta) \tag{1}$$

where θ is the model parameters, and (X, X_T) is a (*sentence, tree*) pair in training set τ, then

$$\begin{aligned} &\log p(X_T|X;\theta) \\ &= \sum_{i=1}^{n} \log p(x_T^i|x_T^1, x_T^2, ...x_T^{i-1}, X;\theta) \end{aligned} \tag{2}$$

where $p(x_T^i|x_T^1, x_T^2, ...x_T^{i-1}, X;\theta)$ is calculated by the decoder.

5 Experimentation

In this section, we introduce the datasets used for evaluation and the baseline methods employed for comparison. We then report the experimental results conducted from different perspectives, and analyze the effectiveness of the proposed model with different factors.

5.1 Settings

We evaluate the effectiveness of our proposed method on a benchmark proposed in Campagnano et al. (2022), which combines several datasets. We specifically choose three datasets (Mohammad et al., 2015; Kim and Klinger, 2018; Bostan et al., 2020) from this benchmark because they are the only ones that encompass annotations for all constituents. We enhance the quality of the original dataset by deleting duplicate or irrelevant examples (e.g., those with all trigger, holder, and target are “null”), while ensuring the trigger is not “null”. The distribution of dataset is shown in Table 1. We choose 90% of the data as the training set, and the remaining 10% of the data as the testing set. Additionally, 10% of the training data is randomly set aside for validation during the training process.

Methods	Emotion	Sentiment	Trigger	Holder	Target	Micro F1.
BERT-CRF	40.3	58.6	19.0	23.3	24.0	33.1
BERT-MRC	42.6	60.3	15.2	22.4	19.3	30.8
TaskFusion	45.7	62.3	-	-	-	-
USSA	41.3	-	17.3	32.6	14.3	-
T5	40.2	59.4	21.0	47.2	32.1	38.2
ChatGPT	22.3	30.7	10.4	21.3	9.6	18.9
LLaMA	42.4	57.6	22.4	43.5	31.3	39.4
Ours	46.3	63.6	30.8	55.0	37.4	46.7

Table 2: Comparison with baselines.

Our proposed method utilizes the generative backbone Flan-T5 (Chung et al., 2022) from HuggingFace's hub. All experiments are conducted on a single NVIDIA Tesla V100 GPU with 32GB memory. The model is trained for 10 epochs in each of the two stages, using the AdamW (Loshchilov and Hutter) optimizer, with a maximum sequence length of 256, a batch size of 16, and a learning rate of 3e-4. And the average results over 5 runs with different random seeds are reported. We adopt F1-score as the metric for all experiments, the micro F1 is employed to evaluate the overall performance.

5.2 Main Results

To better evaluate the effectiveness of our proposed model, we compare its performance against several strong baseline models, which are detailed below,

- **BERT-CRF** (Campagnano et al., 2022) employs bidirectional LSTM layers over BERT representation, then uses CRF to generate token-level labels to determine the span boundary of each semantic constituent.

- **BERT-MRC** (Devlin et al., 2019) is used to predict the start and end indexes of all required semantic constituents.

- **TaskFusion** (Zanwar et al., 2023) is a multi-task framework with a pre-trained language model for emotion-related tasks. We adopt it for the proposed emotion semantic parsing task.

- **USSA** (Zhai et al., 2023) is a joint learning framework, which transfers aspect-based sentiment analysis into a 2D table-filling scheme. We adopt it for the proposed emotion semantic parsing task.

- **T5** (Chung et al., 2022) is a pre-trained language model which naturally suits the sequence-to-sequence task. It also can be considered as a joint learning framework, which generates all the semantic constituents jointly.

- **ChatGPT** (Ouyang et al., 2022) uses dialog-based instructional data in the supervised and RL-based meta-training stages. We use ChatGPT API[1] to generate all required semantic constituents.

- **LLaMA** (Touvron et al., 2023) is a series of large language models trained on trillions of tokens. We use Alpaca-LoRA[2] to fine-tune LLaMA-7B to generate required semantic constituents.

As demonstrated in Table 2, several key observations can be made: 1) Aligning with the current trend favoring generative frameworks, models that adopt a generative paradigm (i.e., T5, LLaMA, and our proposed model) generally outpace models based on a classification paradigm (i.e., BERT-CRF, BERT-MRC, Task Fusion, USSA) in terms of performance. 2) A standout observation is the substantial performance disparity between ChatGPT and all other methodologies. This discrepancy might be attributed to our limited usage of ChatGPT's API, without fine-tuning its internal parameters using insights from our training dataset.

In contrast, our proposed model consistently outperforms prior studies ($p < 0.05$) in all settings. This underscores the effectiveness of our model, employing task decomposition strategies and the abstract semantic tree structure.

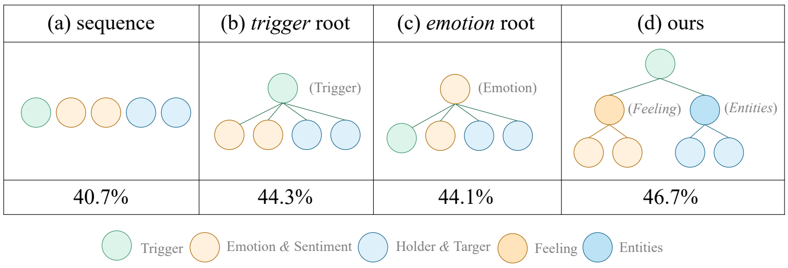

Figure 4: Effect of tree structures.

Tuples	e	s	t	h	a
(e)	42.5	-	-	-	-
(t)	-	-	28.4	-	-
(e, s)	43.5	65.1	-	-	-
(e, t)	45.2	-	29.1	-	-
(e, s, t)	45.7	67.5	30.7	-	-
(t, h)	-	-	29.3	43.1	-
(t, a)	-	-	29.1	-	32.4
(t, h, a)	-	-	29.3	50.0	35.5
(e, t, h, a)	45.9	-	29.4	52.4	35.3
Ours	46.3	63.6	30.8	55.0	37.4

Table 3: Correlations among semantic constituents.

Methods	**Micro F1.**
Ours	46.7
- Mono	39.8
- TriggerPair	40.3
- EntitiesPair	41.8
- FeelingsPair	44.3

Table 4: Impact of decomposition strategies.

5.3 Influence of Semantic Constituents

We then analyze the influence of the semantic constituents and the correlations among them. As shown in Table 3, e denotes the emotion label, s is the sentiment polarity, t, h and a are trigger, holder, and target, separately. In addition, the rows in the table mean different combinations of emotion-related semantic constituents. For example, (e, s) means that we used the proposed model to generate the emotion label and the sentiment polarity jointly.

From the table, it is evident that there is a clear correlation between emotion and sentiment polarity, as the combined performance of (e, s) surpasses that of generating emotion e alone. This underscores the importance of considering sentiment polarity when analyzing and interpreting emotions. Additionally, our findings reveal that the majority of semantic constituents contribute significantly to emotion detection, especially trigger t. For example, the combined performance of (e, t) surpasses that of generating emotion e alone by a large margin. This underscores the importance of these semantic elements in understanding and contextualizing emotions within a given text.

Furthermore, our results highlight a correlation between the holder h and target a with the trigger t, as they often refer to the person or entity experiencing or feeling the emotion triggered by a specific event or situation. This emphasizes the interconnected nature of these semantic constituents and their collective role in shaping emotional expressions.

Notably, we observe that incorporating all semantic constituents leads to the highest performance, highlighting the effectiveness of our proposed model with semantic tree parsing. This underscores the utility and relevance of these semantic constituents in enhancing emotion detection accuracy.

5.4 Effect of Tree Structures

To assess the impact of the abstract semantic tree structure, we evaluated our model's performance using four distinct tree structures in Figure 4.

In our initial assessment, we removed the entire tree structure, treating the semantic constituents as a linear sequence (a). This modification resulted in a substantial decrease in performance, from 46.7% to 40.7%, emphasizing the importance of the tree structure in capturing the intricate relationships among semantic constituents.

Subsequently, we constructed two simplified tree structures, with either *Trigger* (b) or *Emotion* (c) serving as the root node. Interestingly, both of these structures yielded comparable performance results, suggesting that both of them are pivotal components in the emotion parsing task, and their roles are equally significant. Furthermore, we introduced two virtual nodes, labeled as *Feelings* and

[1]https://api.openai.com/v1/chat/completions
[2]https://github.com/tloen/alpaca-lora

Input	Model	Output (Linearized Abstract Semantic Tree)
Jean Valjean was sincere. This sincerity from the very grief that it caused him, rendered inquiries useless, and conferred authority on all that that man had said. Here, for Marius, there was a strange reversal of situations.	T5	[Trigger, *sincerity*, [Feelings, [Emotion, *trust*], [Sentiment, *positive*]], [Entity, [Holder, *him*], [Target, *null*]]]
	Ours	[Trigger, *grief*, [Feelings, [Emotion, *sadness*], [Sentiment, *negative*]], [Entity, [Holder, *him*], [Target, *null*]]]
"Do you think me so very proud?" he suddenly asked. "I think you very strange."	T5	[Trigger, *so very proud*, [Feelings, [Emotion, *joy*], [Sentiment, *positive*]], [Entity, [Holder, *you*], [Target, *me*]]]
	Ours	[Trigger, *suddenly*, [Feelings, [Emotion, *surprise*], [Sentiment, *other*]], [Entity, [Holder, *he*], [Target, *null*]]]
Last year I was wild to marry you, and you wouldn't look at me.	T5	[Trigger, *wouldn't look at me*, [Feelings, [Emotion, *anger*], [Sentiment, *negative*]], [Entity, [Holder, *I*], [Target, *you*]]]
	Ours	[Trigger, *wouldn't look at me*, [Feelings, [Emotion, *disgust*], [Sentiment, *negative*]], [Entity, [Holder, *you*], [Target, *I*]]]

Figure 5: Examples of case study. The text in green means an exact match, while text in red indicates inaccuracies. "Null" means the extracted constituent is missing.

Entities into the tree structure (d). The inclusion of these nodes led to a noticeable improvement in performance, highlighting the efficacy of our proposed tree structures and reinforcing the importance of considering semantic constituents within a structured framework for accurate emotion parsing.

5.5 Impact of Decomposition Strategies

We further investigated the influence of various decomposition strategies, and the results are presented in Table 4, where *Mono* denotes Mono Parsing, *TriggerPair* denotes Trigger-related Pair Parsing, *EntitiesPair* denotes Entities Pair Parsing, and *FeelingsPair* denotes Feelings Pair Parsing. All of these can be found in Section 4.1.

Our analysis reveals that among all the strategies, mono parsing exerts the most substantial impact on model performance, aligning with our initial expectations. Mono parsing can be regarded as the foundational and indivisible task that serves as the basis for other higher-level tasks. Moreover, our findings indicate that all the decomposition strategies employed have a positive influence on the task, emphasizing the importance of breaking down the original task into smaller sub-tasks and the efficacy of reintegrating these sub-tasks to address the main task.

5.6 Case Study

To highlight our model's significance, we've chosen several challenging and illustrative examples shown in Figure 5.

These challenging sentences share the same feature: they express more than one emotion expression, each associated with its own trigger, holder, and target. However, in each case, there is only one overall emotion. The presence of secondary emotions makes it difficult to fully grasp the overall emotional tone of the sentence.

In the first two examples, the mis-detection of emotion expressions by T5 underscores the intricate and nuanced nature of the task, as well as the significance of each element in constructing a precise semantic representation. However, even when the correct emotion expression is identified, modeling the complex relationship between emotion and other semantic components remains challenging. As illustrated in the final example, although T5 correctly identifies the expression and its trigger, its inability to accurately detect the holder and target leads to an incorrect prediction of emotion. This highlights the interconnected nature of these elements and underscores their collective importance in achieving accurate emotion detection.

In contrast, our proposed model, by leveraging task decomposition, demonstrates enhanced efficacy in navigating complexities present in difficult scenarios such as negotiation, emotion shifts, and multiple possible emotion expressions. This not only facilitates accurate detection of individual constituents but also enables a more nuanced understanding of their interrelationships, leading to improved overall performance.

6 Conclusion

In this study, we introduce a novel task called *emotion semantic parsing*, which focuses on analyzing emotions and semantic elements to construct an abstract semantic tree structure. To achieve this, we develop an end-to-end generation model capable of capturing intricate relationships between emotions and various semantic constituents, generating them collaboratively. Furthermore, our proposed model employs template-based prompt learning, allowing for the separate extraction of each semantic constituent and pairs of semantic constituents for individual sub-tasks. Experimental results not only underscore the significance of the introduced task but also showcase the superior performance of our proposed model.

Limitations

Our work focuses on employing a unified framework for emotion semantic parsing. However, one of the limitations is the computational complexity of the proposed model, whose training speed is slower than the classification-based models. Secondly, more tasks could be further explored, including nested emotion parsing, cross-domain and cross-lingual emotion semantic parsing. Finally, a large language model (LLM) can achieve superior performance in nearly all downstream tasks including our task, attributed to their ability to introduce rich knowledge learned during the pre-training stage for our proposed task. For open-source large language models like LLaMA (Touvron et al., 2023), limitations include high computational costs, high-quality data costs, as well as the need to address LLM-related issues such as safety, hallucinations, and ethical concerns, which remain to be resolved in future work.

Acknowledgements

We would like to thank Shufan Fu for her invaluable advice and assistance with the experiments during this work. We also gratefully thank the anonymous reviewers for their insightful and valuable comments. This work is supported by the National Natural Science Foundation of China (No. 62076175, No.61976146), and Jiangsu Innovation Doctor Plan.

References

Fadwa Alrowais, Noha Negm, Majdi Khalid, Nabil Almalki, Radwa Marzouk, Abdullah Mohamed, Mesfer Al Duhayyim, and Amani a Alneil. 2023. Modified earthworm optimization with deep learning assisted emotion recognition for human computer interface. *IEEE Access*, 11:35089–35096.

Nurudin Alvarez-Gonzalez, Andreas Kaltenbrunner, and Vicenç Gómez. 2021. Uncovering the limits of text-based emotion detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2560–2583, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Alexandra Balahur, Jesús M. Hermida, Andrés Montoyo, and Rafael Muñoz. 2011. Emotinet: A knowledge base for emotion detection in text built on the appraisal theories. In *Natural Language Processing and Information Systems*, pages 27–39, Berlin, Heidelberg. Springer Berlin Heidelberg.

Laura Ana Maria Bostan, Evgeny Kim, and Roman Klinger. 2020. GoodNewsEveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1554–1566, Marseille, France. European Language Resources Association.

Cesare Campagnano, Simone Conia, and Roberto Navigli. 2022. SRL4E – Semantic Role Labeling for Emotions: A Unified Evaluation Framework. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4586–4601, Dublin, Ireland. Association for Computational Linguistics.

Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022. Enhanced Multi-Channel Graph Convolutional Network for Aspect Sentiment Triplet Extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2974–2985, Dublin, Ireland. Association for Computational Linguistics.

Xinhong Chen, Qing Li, and Jianping Wang. 2020. A unified sequence labeling model for emotion cause pair extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 208–218, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Kalpana Chowdary, Tu Nguyen, and D. Hemanth. 2021. Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Computing and Applications*, pages 1–18.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, et al. 2022. Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Marwan Dhuheir, Abdullatif Albaseer, Emna Baccour, Aiman Erbad, Mohamed M. Abdallah, and Mounir Hamdi. 2021. Emotion recognition for healthcare surveillance systems using neural networks: A survey. *CoRR*, abs/2107.05989.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020a. ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3161–3170, Online. Association for Computational Linguistics.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020b. End-to-end emotion-cause pair extraction based on sliding window multi-label learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3574–3583, Online. Association for Computational Linguistics.

Fang Fang, Yin Zhou, Shi Ying, and Zhijuan Li. 2023. A study of the ping an health app based on user reviews with sentiment analysis. *International Journal of Environmental Research and Public Health*, 20(2):1591.

Jonas Groschwitz, Shay Cohen, Lucia Donatelli, and Meaghan Fowlie. 2023. AMR parsing is far from solved: GrAPES, the granular AMR parsing evaluation suite. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10728–10752, Singapore. Association for Computational Linguistics.

Xiaojie Gu, Renze Lou, Lin Sun, and Shangxin Li. 2023. Page: A position-aware graph-based model for emotion cause entailment in conversation. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.

Dou Hu, Lingwei Wei, and Xiaoyong Huai. 2021. DialogueCRN: Contextual reasoning networks for emotion recognition in conversations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7042–7052, Online. Association for Computational Linguistics.

Evgeny Kim and Roman Klinger. 2018. Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1345–1359, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Pan Li, Ping Li, and Kai Zhang. 2023. Dual-channel span for aspect sentiment triplet extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 248–261, Singapore. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6818–6825.

Saif M. Mohammad, Xiaodan Zhu, Svetlana Kiritchenko, and Joel Martin. 2015. Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing Management*, 51(4):480–499.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.

Gabriela Pavarini, Tessa Reardon, Anja Hollowell, Vanessa Bennett, Emma Lawrance, Peer Support Young People's Advisory Group Brooks-Hall Ellie Foster-Estwick Ashley Juma Damian Omari Lewis Peter Power Lucy Rogers Maia, Vanessa Pinfold, and Ilina Singh. 2023. Online peer support training to promote adolescents' emotional support skills, mental health and agency during covid-19: Randomised controlled trial and qualitative evaluation. *European Child & Adolescent Psychiatry*, 32(6):1119–1130.

Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4):344–350.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. MISC: A mixed strategy-aware model integrating COMET for emotional support conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 308–319, Dublin, Ireland. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Fanfan Wang, Jianfei Yu, and Rui Xia. 2023. Generative emotion cause triplet extraction in conversations with commonsense knowledge. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 3952–3963, Singapore. 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, Florence, Italy. Association for Computational Linguistics.

Debin Xiao, Rui Xia, and Jianfei Yu. 2023. Emotion cause extraction on social media without human annotation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1455–1468, Toronto, Canada. Association for Computational Linguistics.

Jianfei Yu, Luís Marujo, Jing Jiang, Pradeep Karuturi, and William Brendel. 2018. Improving multi-label emotion classification via sentiment classification with dual attention transfer network. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1097–1102, Brussels, Belgium. Association for Computational Linguistics.

Chaofa Yuan, Chuang Fan, Jianzhu Bao, and Ruifeng Xu. 2020. Emotion-cause pair extraction as sequence labeling based on a novel tagging scheme. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3568–3573, Online. Association for Computational Linguistics.

Sourabh Zanwar, Xiaofei Li, Daniel Wiechmann, Yu Qiao, and Elma Kerz. 2023. What to fuse and how to fuse: Exploring emotion and personality fusion strategies for explainable mental disorder detection. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8926–8940, Toronto, Canada. Association for Computational Linguistics.

Zepeng Zhai, Hao Chen, Ruifan Li, and Xiaojie Wang. 2023. USSA: A unified table filling scheme for structured sentiment analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14340–14353, Toronto, Canada. Association for Computational Linguistics.

Xiaopeng Zheng, Zhiyue Liu, Zizhen Zhang, Zhaoyang Wang, and Jiahai Wang. 2022. UECA-prompt: Universal prompt for emotion cause analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7031–7041, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.