One Unified Model for Diverse Tasks: Emotion Cause Analysis via Self-Promote Cognitive Structure Modeling

Zhaoxin Yu[1,2], Xinglin Xiao[3], Wenji Mao[1,2]*
[1]State Key Laboratory of Multimodal Artificial Intelligence Systems,
Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]Beijing Wenge Technology Co., Ltd
{yuzhaoxin2024,wenji.mao}@ia.ac.cn;xinglin.xiao@wenge.com

Abstract

Emotion cause analysis is a critical topic in natural language processing. Key tasks include emotion cause extraction (ECE), emotion-cause pair extraction (ECPE), social emotion cause identification (SECI) as well as social emotion mining and its cause identification (SEMCI). While current emotion cause analysis methods often focus on task-specific model design, they tend to overlook the underlying common ground across these tasks rooted in cognitive emotion theories, in particular, the *cognitive structure of emotions*. Drawing inspiration from this theory, in this paper, we propose a unified model capable of tackling diverse emotion cause analysis tasks, which constructs the *emotion cognitive structure* through LLM-based in-context learning. To mitigate the hallucination inherent in LLMs, we introduce a self-promote mechanism built on iterative refinement. It dynamically assesses the reliability of substructures based on their cognitive consistency and leverages the more reliable substructures to promote the inconsistent ones. Experimental results on multiple emotion cause analysis tasks ECE, ECPE, SECI, and SEMCI demonstrate the superiority of our unified model over existing SOTA methods and LLM-based baselines.

1 Introduction

With the continuous developments of the Internet and social media, the analysis of emotions and their causes in text has attracted increasing research attention and widely applied to a variety of domains. Previously, multiple emotion cause analysis tasks have been proposed, including emotion cause extraction (ECE) (Lee et al., 2010; Gui et al., 2018), emotion cause pair extraction (ECPE) (Xia and Ding, 2019), social emotion cause identification (SECI) (Xiao et al., 2023a), and social emotion mining and its Cause Identification (SEMCI) (Xiao et al., 2023b). Among them, ECE and ECPE extract the emotions experienced by characters in textual documents, which are usually explicitly expressed by the *author*. In contrast, SECI and SEMCI identify the emotions evoked to the *reader* (i.e., the public) from textual descriptions, which are often implicitly conveyed.

*Corresponding author

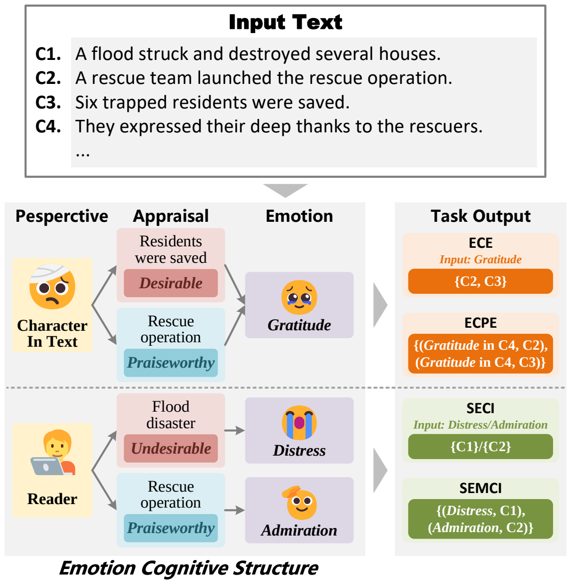

Figure 1: Diverse emotion cause analysis tasks share a common underlying *emotion cognitive structure* and thus can be addressed in a unified manner via cognitive structure modeling.

Existing research on emotion cause analysis has solely focused on specific computational models tailored to individual tasks (Gui et al., 2017; Ding et al., 2020; Xiao et al., 2023a,b), overlooking the underlying common ground rooted in emotion theories. Appraisal theories are well-established and the most influential theoretical models for explaining the antecedents and consequences of human emotions (Lazarus, 1991; Scherer et al., 2001; Sloman et al., 2005; Gratch and Marsella, 2013). They argue that emotions arise from the cognitive pro-

Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 10278–10293

cess of the subjective assessment of personal relationships with the environment, including not only current conditions but also events that lead to this state and future prospects. Appraisal is informed by this cognitive process and maps its characteristics onto a common set of intermediate terms (i.e., appraisal variables).

Among cognitive appraisal theories, the Cognitive Structure of Emotions model (Ortony et al., 1990), proposed by **O**rtony, **C**lore and **C**ollins (i.e., **the OCC model**), is well-developed and the most widely adopted in computational emotion modeling (Smith and Carette, 2022). The OCC model identifies the underlying cognitive structure of different emotion types with elicitation conditions, represented as appraisal dimensions with corresponding values. Specifically, *Desirability*, *Praise/blameworthiness* and *Likelihood* are the three key appraisal dimensions that can lead to certain major emotion types. The *Desirability* dimension is associated with the goal attainments of agents, with *Desirable* and *Undesirable* values; the *Praise/blameworthiness* dimension is associated with certain standards of actions, with *Praiseworthy* and *Blameworthy* values; and the *Likelihood* dimension is associated with the expectation of events, with *Certain* and *Uncertain* values. For example, in Figure 1, past events in the text happened with *Certainty*; the flood disaster is *Undesirable* to the public, leading to *Distress*; the rescue operation is a *Praiseworthy* deed, leading to *Admiration* toward the rescuers. In addition, the combination of these different appraisal dimensions and their values can further elicit compound emotions, such as *Gratitude* in Figure 1, which is elicited by dimensional values *Desirable* and *Praiseworthy* of the character in the text.

Inspired by the OCC model, we take the *emotion cognitive structure* (i.e., *ECS*) as the common cognitive representation and design a unified model to construct the underlying *ECS* via modeling the process of cognitive appraisal, emotion elicitation, and cause analysis, with the *appraisal perspective* (whether from the public, author, or characters in text) being incorporated as an input variable, thereby addressing diverse emotion cause analysis tasks. Previous emotion cause analysis methods typically involved training or fine-tuning discriminative models (Gui et al., 2017; Ding et al., 2020; Xiao et al., 2023a), which primarily captured linguistic features but failed to account for the inherent cognitive structure of emotion.

Recently, large language models (i.e., LLMs) have demonstrated human-like cognitive reasoning capabilities (Rae et al., 2021; Huang and Chang, 2023; Wei et al., 2022b). This progress has positioned LLM-based in-context learning (i.e., ICL) (Brown et al., 2020; Dong et al., 2024; Wei et al., 2022a) as a new paradigm in natural language processing. Therefore, LLMs can be instructed to perform emotion cognitive reasoning (Wang et al., 2023c; Kheiri and Karimi, 2023). However, they still face the challenge of hallucinations (Huang et al., 2024), which can be classified into two types: (1) random errors, and (2) the lack of emotion cognitive capabilities in specific situations. Although some methods improve LLMs' performance on complex reasoning tasks by employing strategies like step-by-step reasoning (Wei et al., 2022b; Kojima et al., 2022) and problem decomposition (Zhou et al., 2023; Wang et al., 2023a), they remain limited by hallucinations caused by randomness. Moreover, other methods improve the reliability of LLMs' outputs via self-refinement (Madaan et al., 2024) or consistency checks (Wang et al., 2023b), yet they do not fundamentally enhance the emotion cognitive capabilities of LLMs.

To tackle these challenges, we propose a **S**elf-**P**romote **E**motion **C**ognitive **S**tructure **(SPECS)**, implemented by LLM-based ICL. Specifically, cause, text and scene-level emotion cognitive reasoning is iteratively executed to update the corresponding substructures and refine the random errors. Furthermore, to better mitigate hallucinations arising from the lack of cognitive capabilities, we propose a self-promote mechanism: first, cognitive consistency across iterations is calculated to assess the reliability of each substructure in *ECS*. Subsequently, since the cognitive capabilities of LLMs in the ICL paradigm largely depend on the few-shot demonstrations, we select demonstrations from high-consistency substructures to promote the low-consistency ones, considering both the cognitive consistency of the candidates and the context relevance across different substructure levels. With this mechanism, LLMs will achieve a promotion of emotion cognitive capabilities from the structures constructed by themselves, thereby further facilitating the refinement of *ECS*.

Our main contributions are as follows:

- Based on the cognitive appraisal theory, we identify the common ground underlying diverse emotion cause analysis tasks ECE,

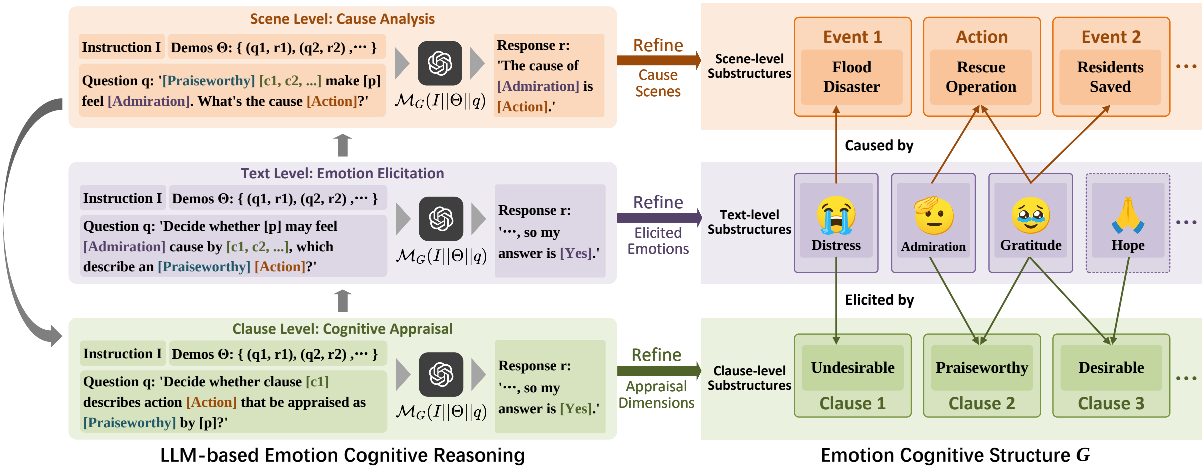

Figure 2: Our proposed emotion cognitive structure consists of a three-level substructure: clause, text, and scene, which respectively model the processes of cognitive appraisal, emotion elicitation, and cause analysis in human emotional cognition.

ECPE, SECI, and SEMCI, and propose a unified model addressing these tasks via cognitive structure modeling.

- We leverage the in-context learning capabilities of LLMs for clause, text, and scene levels emotion cognitive reasoning, and mitigate the hallucination challenge of LLMs through a novel self-promote mechanism.
- We conduct experiments across diverse emotion cause analysis tasks, and the results verify that our unified model significantly outperforms the existing SOTA methods for each individual task as well as LLM-based baselines.

2 Method

In this paper, we aim to design a unified model capable of addressing diverse emotion analysis tasks, including ECE, ECPE, SECI and SEMCI. For ECE and SECI tasks, given the text D and an emotion E, the model outputs a set of cause clauses $\mathcal{C}$. For the ECPE and SEMCI tasks, given the text D only, the model outputs emotion-cause pairs $\{(E, \mathcal{C}_E)\}$.

We model the emotion cognitive structure within the paradigm of LLM-based ICL. Moreover, we propose a self-promote framework to address the hallucination challenge.

2.1 LLM-based Emotion Cognitive Reasoning

The OCC model uses appraisal dimension variables to represent the cognitive appraisal process and establishes a mapping between these variables (or their combinations) and emotions of 22 types. As shown in Table 1, we focus on three key appraisal dimension variables: *Desirability*, *Praise/Blameworthiness*, and *Likelihood*, corresponding to 12 common emotion types. Figure 1 illustrates an example: *Praise* actions that *Certainly* led to *Desirable* events may evoke the emotion of Gratitude.

Appraisal Dimension Variables and Values				Emotion Types
Desirability	Blame/Praise worthiness	Likelihood	Perspective	
Desirable	-	Certain	-	Joy
Undesirable	-	Certain	-	Distress
-	Praise	Certain	Self	Pride
-	Blame	Certain	Self	Shame
-	Praise	Certain	Other	Admiration
-	Blame	Certain	Other	Reproach
Desirable	Praise	Certain	Self	Gratification
Undesirable	Blame	Certain	Self	Remorse
Desirable	Praise	Certain	Other	Gratitude
Undesirable	Blame	Certain	Other	Anger
Desirable	-	Uncertain	-	Hope
Undesirable	-	Uncertain	-	Fear

Table 1: The cognitive association between different combinations of appraisal values and the emotion types they evoke based on the OCC model.

Guided by the OCC model, we designed three fundamental emotion cognitive reasoning operations to construct the emotion cognitive structure (i.e., *ECS*), as shown in Figure 2.

Clause Level: Cognitive Appraisal For each clause, the model respectively appraises the values of three appraisals aforementioned, thereby constructing clause-level substructures within the *ECS*.

Text Level: Emotion Elicitation Applying OCC

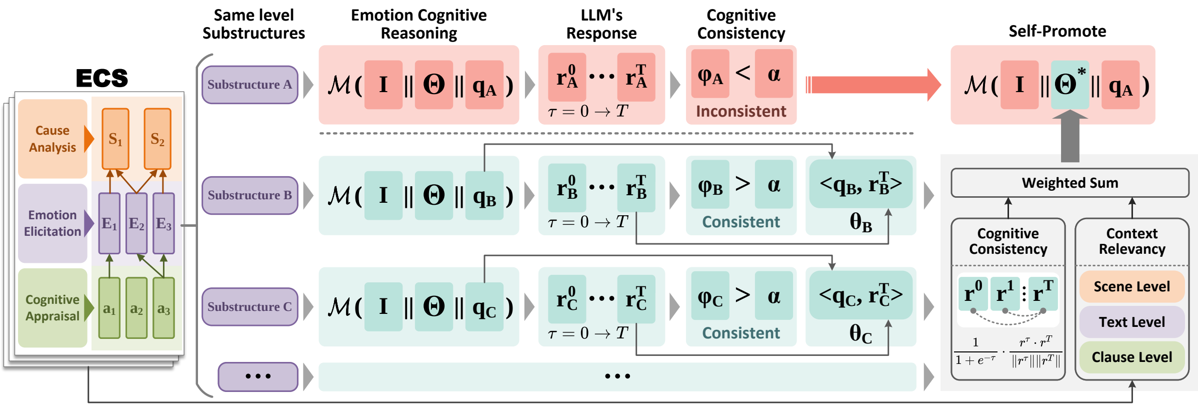

Figure 3: The overall architecture of our self-promote mechanism for cognitively inconsistent substructures.

rules, we compute potential emotions based on the appraisals of clauses and their combinations. The model subsequently evaluates the causal relationships between these appraisals and the corresponding emotions, verifying their presence and constructing text-level substructures within the *ECS*.

Scene Level: Cause Analysis Based on the substructures from the previous two levels (clauses that trigger specific appraisals and the corresponding emotions), the model extracts or generalizes the causes of emotions at the scene level.

The emotion cognitive reasoning of the above process is implemented within the paradigm of LLM-based ICL:

$$r = \mathcal{M}(I||\Theta||q) \tag{1}$$

I, Θ and q represent the system instruction, few-shot demonstrations set, and the input question, respectively. Due to the inherent randomness, LLMs may occasionally produce erroneous reasoning responses. Consequently, *ECS* constructed through a single pipeline may lack full reliability. To address this issue, we propose executing the aforementioned operations iteratively. The previously constructed *ECS* serves as prior knowledge for the current step, and its output modifies the corresponding substructures, enabling the refinement of *ECS*.

2.2 Self-Promote ECS Refinement

Iterative refinement can effectively mitigate hallucinations arising from inherent randomness. However, hallucinations still persist in reasoning cases that exceed the cognitive capabilities of LLMs. This phenomenon parallels human reasoning: under uncertainty or low confidence, humans are unable to provide a consistent and correct answer to repeated questions (Schaeffer and Presser, 2003).

To further enhance the cognitive capabilities of LLMs in the aforementioned situations, we propose a self-promote *ECS* refinement framework, which consists of two stages: self-assessment and self-promote. Figure 3 illustrates the key concept of this framework.

Self-Assessment During Iterative Refinement In this stage, after iteratively performing T_0 iterations of *ECS* refinement with manually designed standard demonstrations, *ECS* of multiple samples are initially constructed. For each substructure g within each *ECS*, we calculate the cognitive consistency ϕ_g across iterative refinements:

$$f_\phi(\{r_g^\tau\}_{\tau=1}^T) = \sum_{\tau=1}^{T-1}\left(\frac{1}{1+e^{-\tau}} \cdot \frac{r^\tau \cdot r^T}{\|r^\tau\|\|r^T\|}\right) \tag{2}$$

Here, $\{r_g^\tau\}_{\tau=1}^T$ denotes the multi-round outputs of emotion cognitive reasoning corresponding to substructure g, which are encoded into fixed-length vectors by Sentence-BERT (Reimers and Gurevych, 2019). We also designed iteration weights $\frac{1}{1+e^{-\tau}}$, which increase as the structure undergoes iterative refinement, assigning greater importance to later iterations. As many studies have shown that the accuracy of LLM-generated content is positively correlated with consistency across iterations (Wang et al., 2023b; Chen et al., 2024; Rabinovich et al., 2023; Xie et al., 2024) , we use cognitive consistency to represent the reliability of each substructure: if $\phi_g > \alpha$ (where α is the consistency threshold), substructure g will be classified as consistent; otherwise, it will be classified as inconsistent.

Self-Promote In-Context Learning In emotion cognitive reasoning, the selection of few-shot demonstrations for ICL significantly impacts the cognitive capabilities of LLMs. Therefore, we pro-

Algorithm 1: Process of Unified Model

Input: Dataset $\mathcal{D} = \{D_k\}_{k=1}^{N_\mathcal{D}}$
Output: ECE/SECI task: cause clauses $\mathcal{C}_k$ for each D_k; ECPE/SEMCI task: emotion-clause pairs $\{(E, \mathcal{C}_E)\}_k$ for each D_k

// Initialization of ECS
foreach *iteration* τ *in* $\{1, \cdots, T_0\}$ **do**
  foreach k *in* $\{1, \cdots, N_\mathcal{D}\}$ **do**
    Reasoning: perform emotion cognitive reasoning $r_g^\tau \leftarrow \mathcal{M}_{G_k}(I||\Theta||q_g)$ at the corresponding level for each substructure g in G_k;
    Refinement: update G_k by r_g^τ;
// Self-Promote ECS Refinement
foreach τ' *in* $\{T_0 + 1, \cdots, T_1\}$ **do**
  // Self-Assessment
  Consistent substructures $\mathcal{K} \leftarrow \{\}$;
  foreach k *in* $\{1, \cdots, N_\mathcal{D}\}$ **do**
    foreach *substructure* g *in* G_k **do**
      Cognitive consistency: $\phi_g \leftarrow f_\phi(\{r_g^\tau\}_{\tau=1}^{\tau'-1})$;
      if $\phi_g > \alpha$ **then**
        $\mathcal{K} \leftarrow \mathcal{K} \cup \{g\}$;
        $r_g^{\tau'} \leftarrow r_g^{\tau'-1}$;
  // Self-Promote
  foreach k *in* $\{1, \cdots, N_\mathcal{D}\}$ **do**
    foreach *substructure* g *in* G_k **do**
      if g *not in* $\mathcal{K}$ **then**
        Select Θ_g^* from $\mathcal{K}$;
        $r_g^\tau \leftarrow \mathcal{M}_{G_k}(I||\Theta_g^*||q_g)$;
// Answer Extraction
Answers set $\mathcal{A} \leftarrow \{\}$;
foreach k *in* $\{1, \cdots, N_\mathcal{D}\}$ **do**
  if *Task = ECE/SECI* **then**
    Extract cause clauses $\mathcal{C}_k$ from G_k;
    $\mathcal{A} \leftarrow \mathcal{A} \cup \{\mathcal{C}_k\}$;
  else if *Task = ECPE/SECMI* **then**
    Extract emotion-clause pairs $\{(E, \mathcal{C}_E), \cdots\}_k$ from G_k;
    $\mathcal{A} \leftarrow \mathcal{A} \cup \{\{(E, \mathcal{C}_E), \cdots\}_k\}$;
return $\mathcal{A}$

pose extracting demonstrations from cognitively consistent substructures to promote the inconsistent ones. Cognitive consistency and context relevancy at different levels will be simultaneously considered as factors in the selection of demonstrations. Specifically, we compute the score by performing a weighted summation of the cognitive consistency ϕ_j of the consistent candidate substructure g_j and the context relevancy $\rho_{(i,j)}$ with the target substructure g_i:

$$SCORE = \delta_0 \cdot \phi_j + \sum_{l=1}^{3} (\delta_l \cdot \rho_{(i,j)}^l) \quad (3)$$

Here, $\delta_0, \cdots, \delta_3$ represent the weights of cognitive consistency and the context relevancy at the three levels of *ECS*: clause, text and scene, respectively. The cosine similarity is used to measure the relevance of the textual context:

$$\rho_{(i,j)}^l = \frac{t_i^l \cdot t_j^l}{||t_i^l|| ||t_j^l||} \quad (4)$$

Here, t_i^l and t_j^l represent the embedding of textual context at level l.

In the subsequent *ECS* iterative refinement process, for each inconsistent substructure, a set of top-ranked positive and negative demonstrations is selected from all consistent substructures at the same level from other *ECS*. This mechanism is executed dynamically to continuously update each *ECS*'s consistency status, allowing the LLM to learn from its own high-consistency reasoning and achieve self-promote in emotion cognition.

2.3 Unified ECS Modeling

Formally, Algorithm 1 outlines the complete process of our proposed model. The model conducts explicit emotion cognitive reasoning within the LLM-based ICL paradigm, denoted as $\mathcal{M}_{G_k}(I||\Theta||q_g)$. Here $\Theta = \{\theta_1, \theta_2, \cdots\}$ represents the set of few-shot demonstrations, and $\mathcal{M}_{G_k}$ denotes the current *ECS* G_k serving as prior knowledge for emotion cognitive reasoning before the query q_g.

For ECE/SECI tasks, the clause-level substructure corresponding to the given emotion E is extracted as cause clauses. For ECPE/SEMCI tasks, in addition to the clauses, emotion expressions at scene or text level is also extracted.

3 Experiments

3.1 Dataset and Metrics

We evaluated the performance of our proposed method on the ECE, ECPE, SECI, and SEMCI tasks. Experiments were conducted on two publicly available emotion cause analysis datasets. For ECE and ECPE tasks, we conduct experiments based on the **ECPE Chinese dataset**[1] (Xia and Ding,

[1] https://github.com/NUSTM/ECPE

2019). The dataset is built upon the classic benchmark ECE corpus for the ECE task and includes a series of Chinese city news from NEWS SINA, along with annotations for emotion categories and emotional cause clauses. For SECI and SEMCI tasks, we conduct experiments based on the **SECI dataset**[2] (Xiao et al., 2023a). The dataset contains a series of Chinese online news documents, which evoke six social emotion types, along with annotations for the cause clauses corresponding to specific emotions. The details of datasets will be provided in Appendix C.

The precision (**P**), recall (**R**) and F1 score (**F1**) defined in (Gui et al., 2018; Xia and Ding, 2019; Xiao et al., 2023a,b) are used to evaluate the performance of the four tasks.

3.2 Baseline Methods

We designed comparative experiments to validate the superiority of the proposed self-promote emotion cognitive structure (SPECS). Firstly, we adopted several high-performing supervised discriminative models in individual tasks as baselines:

ECE: RTHN (Xia et al., 2019), FSS-GCN (Hu et al., 2021b), EF-BHA (Hu et al., 2021a) and UECA-Prompt (Zheng et al., 2022).

ECPE: UECA-Prompt (Zheng et al., 2022), ECPE-MTL (Li et al., 2023), CD-MRC (Cheng et al., 2023), MV-SHIF (Yang et al., 2024) and MGCL (Yu et al., 2024) .

SECI: RTHN (Xia et al., 2019), FSS-GCN (Hu et al., 2021b), BERT-encoded MLP (Devlin et al., 2019) and CogEES (Xiao et al., 2023a).

SEMCI: BERT-encoded MLP (Devlin et al., 2019) and JointPSEC (Xiao et al., 2023b).

Likewise, we evaluated the performance of LLM-based in-context learning methods, including Standard ICL (Ouyang et al., 2022), Chain-of-Thought (Kojima et al., 2022), KATE (Liu et al., 2021), Self-Refine (Madaan et al., 2024), Self-Consistency (Wang et al., 2023b) and COSP (Wan et al., 2023). Qwen2-7B[3], an open-source model and GPT-4o[4], a closed-source model, were selected as the LLMs.

3.3 Main Results

Comparison with supervised discriminative models As shown in Table 2, compared to supervised discriminative models, our SPECS model achieves superior overall performance, with F1-score improvements of 2.87%, 4.78%, 3.26% and 7.78% on the ECE, ECPE, SECI, SEMCI tasks, respectively, over SOTA methods. This improvement is primarily attributed to a significant increase in Precision, which surpasses the SOTA methods by 4.93%, 4.93%, 4.32% and 6.37% on the four tasks, respectively. This is because modeling the underlying emotion cognitive structure enables the capture of deeper cognitive relationships beyond linguistic features, which are solely the focus of supervised discriminative models.

Tasks	Methods	P	R	F1
ECE	RTHN†	76.97	76.62	76.77
	FSS-GCN	78.05	76.13	77.08
	EF-BHA◇†	79.38	78.08	78.68
	UECA-Prompt◇	<u>82.67</u>	<u>84.33</u>	<u>83.49</u>
	SPECS$_{\text{(w/ Qwen2-7B)}}$	86.17	83.88	85.01
	SPECS$_{\text{(w/ GPT-4o)}}$	**87.60**	**85.16**	**86.36**
ECPE	UECA-Prompt◇	72.19	78.04	75.00
	ECPE-MTL◇	75.61	75.04	75.32
	CD-MRC◇	82.53	77.60	79.99
	MV-SHIF†	80.80	78.40	79.60
	MGCL†	<u>83.41</u>	<u>80.13</u>	<u>81.66</u>
	SPECS$_{\text{(w/ Qwen2-7B)}}$	85.16	81.02	83.03
	SPECS$_{\text{(w/ GPT-4o)}}$	**86.30**	**83.71**	**84.99**
SECI	RTHN†	65.42	63.03	64.20
	FSS-GCN†	65.79	64.10	64.93
	BERT+MLP◇	75.47	76.62	76.04
	CogEES◇†	<u>80.41</u>	<u>80.13</u>	<u>80.23</u>
	SPECS$_{\text{(w/ Qwen2-7B)}}$	82.40	81.05	81.72
	SPECS$_{\text{(w/ GPT-4o)}}$	**84.73**	**82.29**	**83.49**
SEMCI	BERT+MLP◇	<u>70.34</u>	<u>69.28</u>	<u>69.81</u>
	JointPSEC†	68.02	67.70	67.86
	SPECS$_{\text{(w/ Qwen2-7B)}}$	74.94	71.28	73.06
	SPECS$_{\text{(w/ GPT-4o)}}$	**76.71**	**74.59**	**75.64**

Table 2: Comparison between our methods and the baselines based on supervised discriminative models, with the best results highlighted in bold and the best results of baselines underlined. ◇ indicates the method is bert-based and † indicates the results are reported in the original paper. Our proposed SPECS method is highlighted in blue in the table.

Comparison with LLM-based ICL frameworks We also evaluated the performance of LLM-based ICL frameworks on these four tasks. The experimental results show that standard ICL under performs supervised discriminative models. This is primarily due to LLM hallucinations, where many

[2] https://github.com/xxllll/social-emotion-cause-identification
[3] https://huggingface.co/Qwen/Qwen2-7B-Instruct
[4] https://platform.openai.com/docs/models/gpt-4o

LLM	ICL Methods	Task: ECE			Task: ECPE			Task: SECI			Task: SEMCI		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1
Qwen2-7B	Standard ICL	72.58	71.24	71.90	65.03	69.44	67.16	65.52	77.58	71.04	62.83	64.50	63.65
	COT	77.73	76.59	77.16	74.32	71.28	72.77	69.78	75.58	72.56	65.65	63.09	64.34
	KATE	81.02	77.62	79.28	79.04	70.80	74.69	79.59	76.13	77.82	72.49	67.53	70.98
	Self-Refine	77.81	80.97	79.36	73.39	76.48	74.90	77.21	78.44	77.82	69.43	68.00	68.71
	Self-Consistency	80.27	82.04	81.15	79.79	78.91	79.35	80.07	77.41	78.72	68.65	67.79	68.22
	COSP	82.69	83.07	82.88	80.30	80.93	80.61	78.64	79.19	78.91	70.47	68.83	69.64
	SPECS	**86.17**	**83.88**	**85.01**	**85.16**	**81.02**	**83.03**	**82.40**	**81.05**	**81.72**	**74.94**	**71.28**	**73.06**
GPT-4o	Standard ICL	77.49	77.93	77.71	74.05	77.91	75.93	72.10	75.33	73.68	67.29	68.54	68.56
	COT	76.58	78.22	77.39	76.39	73.47	74.90	76.00	76.24	76.12	70.28	71.51	70.89
	KATE	83.80	80.07	81.89	80.71	78.94	79.82	80.57	77.39	78.95	71.00	69.94	70.47
	Self-Refine	80.89	84.73	82.77	77.58	**84.12**	80.72	79.02	80.48	79.74	73.44	70.81	72.10
	Self-Consistency	82.96	82.60	82.78	82.17	81.43	81.80	82.83	78.77	80.75	69.76	70.33	70.04
	COSP	84.21	82.10	83.14	82.51	81.77	82.14	82.69	80.25	81.45	74.32	72.40	73.35
	SPECS	**87.60**	**85.16**	**86.36**	**86.30**	83.71	**84.99**	**84.73**	**82.29**	**83.49**	**76.71**	**74.59**	**75.64**

Table 3: Comparison between our methods and the LLM-based ICL frameworks, with the best results highlighted in bold and the second-best results underlined. Our proposed SPECS method is highlighted in blue in the table.

clauses without cognitive relationships to the emotion are mistakenly identified as causes in complex scenarios requiring multi-step emotion reasoning, providing seemingly 'plausible' explanations. Prompting strategies significantly improve LLM performance. However, these methods still require manually designed prompts for each emotion type, limiting the use of LLMs' general reasoning capabilities to create a unified model. Moreover, hallucinations remain a major issue for reasoning cases beyond LLMs' cognitive capabilities. Our proposed SPECS method effectively addresses the aforementioned issues. As shown in Table 3, SPECS based on Qwen2-7B outperforms all the best-performing baselines, achieving F1-score improvements of 2.13%, 2.42%, 2.81%, and 3.42% across the four tasks, respectively. Likewise, SPECS based on GPT-4o achieves F1-score improvements of 3.22%, 2.85%, 2.04%, and 2.29% over the best baselines across the four tasks, respectively.

3.4 Ablation Study

Ablation Study of *ECS* Refinement To further validate the contribution of our proposed *ECS* refinement, we conducted an ablation study. The refinement of the three levels of substructures—cognitive appraisal (clause-level), emotion elicitation (text-level), and cause analysis (scene-level) was individually removed and evaluated across the four tasks. Table 4 illustrates the experimental results with Qwen2-7B. Removing the refinement of any level resulted in a performance decline. In the ECE and SECI tasks, cause analysis had the greatest impact, while emotion elicitation was more critical in the ECPE and SEMCI tasks, where emotion types are unknown. Additionally, cognitive appraisal had a stronger influence on SECI and SEMCI tasks, as social emotions and their causes are often implicit, requiring inference through cognitive appraisal.

Ablation Study of Self-Promote Mechanism We validated the effectiveness of the proposed self-promote mechanism. Figure 4 illustrates the model's performance and the consistent rate $\frac{N_{consistent}}{N_{substructures}}$, where $N_{consistent}$ denotes the number of cognitively consistent substructures and $N_{substructures}$ denotes the total number of substructures. The vertical dashed line in the figure indicates the introduction of the self-promote mechanism at the 5-th iteration, which significantly improves both the model's overall performance and the consistency rate of the emotion cognitive structure. Furthermore, the experimental results reveal that model performance and consistency rate exhibit a correlated pattern, supporting our hypothesis that the higher the cognitive consistency of the substructures, the greater their reliability.

3.5 Analysis of Consistency Threshold

To explore the the impact of the consistency threshold α in our model, various α values were configured to evaluate the model's performance across the four tasks. As shown in Figure 5, in the ECE and ECPE tasks, the model achieves optimal perfor-

Methods	Task: ECE			Task: ECPE			Task: SECI			Task: SEMCI		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
SPECS	**86.17**	**83.88**	**85.01**	**85.16**	**81.02**	**83.03**	**82.40**	**81.05**	**81.72**	**74.94**	**71.28**	**73.06**
w/o Cognitive Analysis	85.08	83.61	84.34	84.25	80.58	82.37	79.60	80.27	79.93	71.65	69.42	70.52
w/o Emotion Elicitation	-	-	-	74.26	80.04	77.04	-	-	-	68.43	68.29	68.36
w/o Cause Analysis	79.38	79.52	79.45	78.20	78.67	78.43	77.84	77.52	77.68	71.49	70.08	70.78
w/o All	77.03	78.14	77.58	72.56	76.66	74.55	75.54	77.20	76.36	65.43	67.96	66.67

Table 4: Results of ablation study, with the refinement of substructures at three levels: Cognitive Appraisal (clause-level), Emotion Elicitation (text-level), and Cause Analysis (scene-level) being removed from the SPECS model respectively.

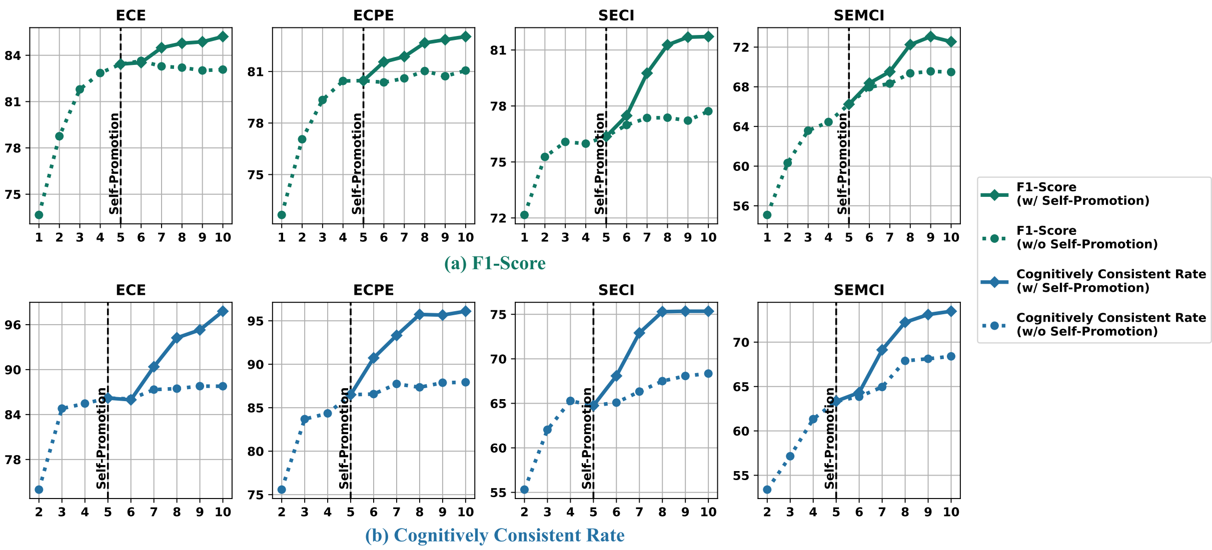

Figure 4: The (a) F1-score and (b) the cognitively consistent rate of substructures across the four tasks through iterations. The vertical dashed line at the fifth iteration indicates the introduction of the self-promote mechanism.

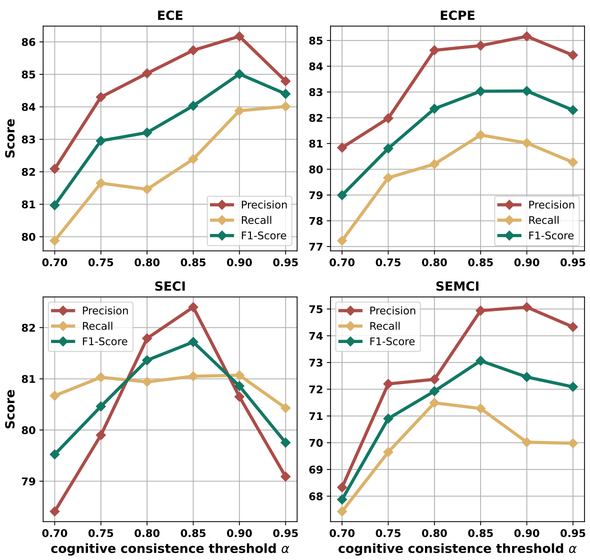

Figure 5: Our model's performance with different consistency threshold α.

mance at $\alpha = 0.9$, while in the SECI and SEMCI tasks, optimal performance is attained at $\alpha = 0.85$. Overall, both the precision and recall of the model exhibit an initial increase followed by a decrease as α increases. On one hand, when α is too large, certain substructures that have already been correctly refined through iterations may be incorrectly classified as cognitively inconsistent, leading to misguidance from other erroneous substructures in the self-promote process. On the other hand, when α is too small, cognitively inconsistent structures cannot be effectively filtered out, resulting in insufficient refinement of the *ECS*.

4 Related Work

Cognitive appraisal theories In cognitive psychology, the causes and effects of emotions have been extensively studied through cognitive appraisal theories (Ortony et al., 1990; Scherer et al., 2001; Sloman et al., 2005; Gratch and Marsella, 2013). These theories argue that emotions arise from the subjective assessment of personal relationships with the environment, including not only current conditions, but also events that lead to this state and future prospects (Gratch et al., 2006). Appraisal itself is influenced and guided by cognitive processes, which map the characteristics of these processes into a common set of intermediate terms (i.e., appraisal variables). Among cognitive appraisal theories, the cognitive structure of emotions model (i.e., the OCC model) proposed by (Ortony et al., 1990) is one of the most well-developed psychological emotion models and widely adopted in computational modeling of emotions (Smith and Carette, 2022). The OCC model identifies the underlying cognitive structure of 22 emotion types: Well-being (e.g. joy, distress), Prospect-based (e.g. hope, fear), Attribution (e.g. admiration, reproach), Well-being/attribution compound (e.g. gratitude, anger), Fortunes-of-others and Attraction (e.g. love, hate). The OCC model has been widely applied to various computational tasks, including character modeling (Klinkert and Clark, 2021), robot-human communication (Olgun et al., 2018), and text mining (Xiao et al., 2023a).

Supervised discriminative models for emotion cause analysis Emotion cause analysis has garnered significant research attention in recent years. Supervised discriminative models focus on the linguistic associations between emotion expression and their causes. Notable research approaches include feature extraction encoders (Xia et al., 2019; Hu et al., 2021a), incorporating graph neural networks (Wei et al., 2020; Hu et al., 2021b; Xiao et al., 2023a), multi-task learning (Li et al., 2021, 2023; Xiao et al., 2023b) and query-aware method (Cheng et al., 2023; Diao et al., 2020). Additionally, in recent years, prompt tuning have demonstrated remarkable performance, such as (Zheng et al., 2022; Zhou et al., 2022; Gu et al., 2024). A common limitation of the aforementioned methods is their focus solely on linguistic-level features, lacking the ability of deeper emotion cognition and interpretability.

LLM-based in-context learning frameworks In recent years, numerous studies have also explored various mechanisms to enhance the overall performance of LLMs, including Chain-of-thought (Wei et al., 2022b), problem decomposition (Zhou et al., 2023; Wang et al., 2023a), demonstrations selection (Rubin et al., 2022; Liu et al., 2021), self-consistency (Wang et al., 2023b) and self-refine (Madaan et al., 2024). Similar to this work, some existing methods (Rubin et al., 2022; Liu et al., 2021; Wan et al., 2023) enhance the LLM's ICL capabilities by demonstrations selection. However, when directly applied to emotion cause analysis tasks, the aforementioned methods struggle with hallucination issues, as they do not fundamentally derive the emotion cognitive reasoning.

5 Conclusion

This paper presents a unified model capable of diverse emotion cause analysis tasks. Our model employs the LLM-based in-context learning to iteratively construct the underlying emotion cognitive structure, which are often overlooked by existing methods. To mitigate the hallucination problem in LLMs, we designed a self-promote mechanism, which enhances LLMs' emotion cognitive capability for cognitively inconsistent reasoning cases, without requiring additional external knowledge or training. Experimental results show that our method outperforms strong baselines across all four tasks, including supervised discriminative models and LLM-based ICL frameworks. Ablation studies further validate the effectiveness of each component in our model.

Limitations

Our approach's limitation lies in the simplification of the OCC model. Specifically, we focused on three appraisals—Desire, Praise/Blameworthiness, and Likelihood—to describe the cognitive appraisal process, enabling the analysis of cause for 12 emotion types. However, the comprehensive OCC model identifies the cognitive structure underlying 22 emotion types, which allows for a more detailed representation of human emotional expression. Expanding our model to incorporate the complete set of emotion types would enable more nuanced and sophisticated emotion cause analysis.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grants #72293575, #72434005, #72225011 and #62206287. We thank the anonymous reviewers for the valuable comments.

References

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda

Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Xinyun Chen, Renat Aksitov, Uri Alon, Jie Ren, Kefan Xiao, Pengcheng Yin, Sushant Prakash, Charles Sutton, Xuezhi Wang, and Denny Zhou. 2024. Universal self-consistency for large language models. In *ICML 2024 Workshop on In-Context Learning*.

Zifeng Cheng, Zhiwei Jiang, Yafeng Yin, Cong Wang, Shiping Ge, and Qing Gu. 2023. A consistent dual-mrc framework for emotion-cause pair extraction. *ACM Transactions on Information Systems*, 41(4):1–27.

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*, page 4171–4186.

Yufeng Diao, Hongfei Lin, Liang Yang, Xiaochao Fan, Yonghe Chu, Di Wu, Kan Xu, and Bo Xu. 2020. Multi-granularity bidirectional attention stream machine comprehension method for emotion cause extraction. *Neural Computing and Applications*, 32:8401–8413.

Zixiang Ding, Rui Xia, and Jianfei Yu. 2020. 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.

Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Jingyuan Ma, Rui Li, Heming Xia, Jingjing Xu, Zhiyong Wu, Baobao Chang, et al. 2024. A survey on in-context learning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1107–1128.

Jonathan Gratch, Wenji Mao, and Stacy Marsella. 2006. Modeling social emotions and social attributions. *Cognition and Multi-Agent Interaction*, pages 219–251.

Jonathan Gratch and Stacy Marsella. 2013. *Social emotions in nature and artifact*. Oxford University Press.

Xue Gu, Zhihan Zhou, Ziyao Meng, Jian Li, Tiago Gomes, Adriano Tavares, and Hao Xu. 2024. EmoPrompt-ECPE: Emotion knowledge-aware prompt-tuning for emotion-cause pair extraction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pages 5678–5688.

Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu, and Jiachen Du. 2017. A question answering approach for emotion cause extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1593–1602.

Lin Gui, Ruifeng Xu, Dongyin Wu, Qin Lu, and Yu Zhou. 2018. Event-driven emotion cause extraction with corpus construction. In *Social Media Content Analysis: Natural Language Processing and Beyond*, pages 145–160. World Scientific.

Guimin Hu, Guangming Lu, and Yi Zhao. 2021a. Bidirectional hierarchical attention networks based on document-level context for emotion cause extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 558–568.

Guimin Hu, Guangming Lu, and Yi Zhao. 2021b. FSS-GCN: A graph convolutional networks with fusion of semantic and structure for emotion cause analysis. *Knowledge-Based Systems*, 212:106584.

Jie Huang and Kevin Chen-Chuan Chang. 2023. Towards reasoning in large language models: A survey. In *Proceedings of the 61st Annual Meeting Of The Association For Computational Linguistics*.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2024. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*.

Kiana Kheiri and Hamid Karimi. 2023. SentimentGPT: Exploiting GPT for advanced sentiment analysis and its departure from current machine learning. *arXiv preprint arXiv:2307.10234*.

Lawrence J Klinkert and Corey Clark. 2021. Artificial psychosocial framework for affective non-player characters. In *Advances in Artificial Intelligence and Applied Cognitive Computing: Proceedings from ICAI'20 and ACC'20*, pages 695–714. Springer.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.

Richard S Lazarus. 1991. *Emotion and adaptation*. Oxford University Press.

Sophia Yat Mei Lee, Ying Chen, Shoushan Li, Chu-Ren Huang, et al. 2010. Emotion cause events: Corpus construction and analysis. In *Proceedings of the 2010 International Conference on Language Resources and Evaluation*.

Chenbing Li, Jie Hu, Tianrui Li, Shengdong Du, and Fei Teng. 2023. An effective multi-task learning model for end-to-end emotion-cause pair extraction. *Applied Intelligence*, 53(3):3519–3529.

Xiangju Li, Shi Feng, Yifei Zhang, and Daling Wang. 2021. Multi-level emotion cause analysis by multi-head attention based multi-task learning. In *Proceedings of the 2021 China National Conference on Chinese Computational Linguistics*, pages 77–93. Springer.

Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. 2021. What makes good in-context examples for GPT-3? *arXiv preprint arXiv:2101.06804*.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. 2024. Self-refine: Iterative refinement with self-feedback. *Advances in Neural Information Processing Systems*, 36.

Saif M Mohammad. 2022. Ethics sheet for automatic emotion recognition and sentiment analysis. *Computational Linguistics*, 48(2):239–278.

Markellos S Nomikos, Edward Opton Jr, and James R Averill. 1968. Surprise versus suspense in the production of stress reaction. *Journal of personality and social psychology*, 8.

Zehra Nur Olgun, YuJung Chae, and ChangHwan Kim. 2018. A system to generate robot emotional reaction for robot-human communication. In *Proceedings of the 15th International Conference on Ubiquitous Robots*, pages 383–387.

Andrew Ortony, Gerald L Clore, and Allan Collins. 1990. *The cognitive structure of emotions*. Cambridge university press.

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.

Ella Rabinovich, Samuel Ackerman, Orna Raz, Eitan Farchi, and Ateret Anaby Tavor. 2023. Predicting question-answering performance of large language models through semantic consistency. In *Proceedings of the Third Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*, pages 138–154.

Jack W Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, et al. 2021. Scaling language models: Methods, analysis & insights from training gopher. *arXiv preprint arXiv:2112.11446*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, page 3982–3992.

Ohad Rubin, Jonathan Herzig, and Jonathan Berant. 2022. Learning to retrieve prompts for in-context learning. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2655–2671.

Nora Cate Schaeffer and Stanley Presser. 2003. The science of asking questions. *Annual review of sociology*, 29(1):65–88.

Klaus R Scherer. 1984. *On the nature and function of emotion: a component process approach. Approaches to emotion*. NJ: Erlbaum, Hillsdale, kr scherer and p. ekman (eds.) edition.

Klaus R Scherer, Angela Schorr, and Tom Johnstone. 2001. *Appraisal processes in emotion: Theory, methods, research*. Oxford University Press.

Aaron Sloman, Ron Chrisley, and Matthias Scheutz. 2005. The architectural basis of affective states and processes. *Who Needs Emotions? The Brain Meets the Robot*.

Geneva M Smith and Jacques Carette. 2022. What lies beneath—a survey of affective theory use in computational models of emotion. *IEEE Transactions on Affective Computing*, 13(4):1793–1812.

Xingchen Wan, Ruoxi Sun, Hanjun Dai, Sercan Arik, and Tomas Pfister. 2023. Better zero-shot reasoning with self-adaptive prompting. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 3493–3514.

Lei Wang, Wanyu Xu, Yihuai Lan, Zhiqiang Hu, Yunshi Lan, Roy Ka-Wei Lee, and Ee-Peng Lim. 2023a. Plan-and-solve prompting: Improving zero-shot chain-of-thought reasoning by large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2609–2634.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023b. Self-consistency improves chain of thought reasoning in language models. In *Proceedings of the Eleventh International Conference on Learning Representations*.

Zengzhi Wang, Qiming Xie, Yi Feng, Zixiang Ding, Zinong Yang, and Rui Xia. 2023c. Is ChatGPT a good sentiment analyzer? a preliminary study. *arXiv preprint arXiv:2304.04339*.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022a. Emergent abilities of large language models. *Transactions on Machine Learning Research*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022b. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Penghui Wei, Jiahao Zhao, and Wenji Mao. 2020. Effective inter-clause modeling for end-to-end emotion-cause pair extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3171–3181.

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.

Rui Xia, Mengran Zhang, and Zixiang Ding. 2019. RTHN: A rnn-transformer hierarchical network for emotion cause extraction. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*.

Xinglin Xiao, Wenji Mao, Ying Sun, and Daniel Zeng. 2023a. A cognitive emotion model enhanced sequential method for social emotion cause identification. *Information Processing & Management*, 60(3):103305.

Xinglin Xiao, Yuan Tian, Yin Luo, and Wenji Mao. 2023b. A cognitive knowledge enriched joint framework for social emotion and cause mining. In *Proceedings of the 2023 International Conference on Knowledge Science, Engineering and Management*, pages 396–405. Springer.

Zhihui Xie, Jizhou Guo, Tong Yu, and Shuai Li. 2024. Calibrating reasoning in language models with internal consistency. *arXiv preprint arXiv:2405.18711*.

Cheng Yang, Hua Zhang, Bi Chen, Bo Jiang, and Ye Wang. 2024. MV-SHIF: Multi-view symmetric hypothesis inference fusion network for emotion-cause pair extraction in documents. *Neural Networks*, 175:106283.

Yang Yu, Xin Lin, Changqun Li, Shizhou Huang, and Liang He. 2024. MGCL: Multi-granularity clue learning for emotion-cause pair extraction via cross-grained knowledge distillation. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 1897–1907.

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.

Changzhi Zhou, Dandan Song, Jing Xu, and Zhijing Wu. 2022. A multi-turn machine reading comprehension framework with rethink mechanism for emotion-cause pair extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6726–6735.

Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc V Le, et al. 2023. Least-to-most prompting enables complex reasoning in large language models. In *Proceedings of the Eleventh International Conference on Learning Representations*.

A Ethics Considerations

Since this work involves the fields of emotion and social computing, it is necessary to discuss the potential ethical issues. Below, we discuss these issues in four aspects: task design, data usage, implications for privacy, and implications for social, following the Ethics Sheet for Sentiment Analysis (Mohammad, 2022).

Task Design The emotion cause analysis tasks involved in our work are essentially natural language processing tasks. Our goal is solely to infer potential information embedded in the text, and it should not be interpreted as an attempt to predict an individual's emotional state.

Data Usage We did not construct or propose any new datasets. All the data used in this work come from publicly available datasets (Xia and Ding, 2019; Xiao et al., 2023a), whose sources are also publicly accessible. Furthermore, all personal information related to the samples has been anonymized.

Implications for Privacy Consider that people might not want their emotions to be inferred. The used data in our work is fixed and we do not engage in continuously gathering emotion-related information. We firmly oppose using the models discussed in this work for any applications that may infringe on personal privacy.

Implications for Social We have neither constructed nor applied any large-scale emotion detection systems, and we firmly oppose using the models discussed in this work for any applications that could lead to negative societal impacts.

B Supplementary Experiment

B.1 Validation of the Correlation Between Cognitive Consistency and the Cognitive Structure Correctness

Our proposed self-promotion mechanism assumes that the reliability of the emotion cognitive substructure correlates positively with its cognitive consistency across iterations. Experimental results in Figure 4 support this hypothesis. To further validate it, we conducted supplementary experiments. Without the self-promotion mechanism, we performed 5 iterations of ECS refinement, calculating the cognitive consistency of each sample's substructures. We then averaged these results to represent the sample's overall cognitive consistency and statistically analyzed the consistency scores

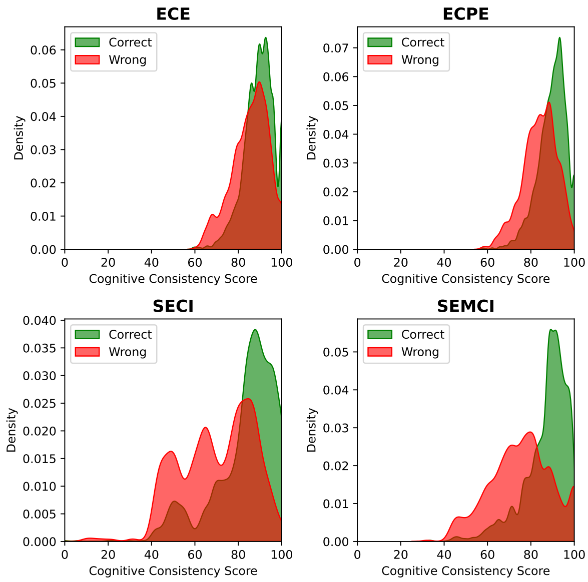

Figure 6: Density distribution of the average cognitive consistency within the emotion cognitive substructures for correctly and incorrectly predicted emotion causes in four tasks

for samples with correct versus incorrect emotion-cause predictions. As shown in Figure 6, the samples with correct predictions tend to exhibit higher cognitive consistency in four tasks, which further validates the aforementioned hypothesis.

B.2 Evaluation with Various LLMs

To validate the robustness of our proposed method across multiple LLMs, we conducted experiments using various LLMs of different sizes beyond Qwen2-7B and GPT-4o. We selected five open-source LLMs of varying sizes[5]: Qwen2-1.5B[6], phi3.5-3.8B[7], LLaMA-3-8B[8], Yi-1.5-34B[9], and Qwen2-72B[10], along with GPT-3.5 Turbo[11], for experiments. All experiments with open-source LLMs are run on the machine containing 4 pieces of Tesla V100 (32GB) GPUs.

As shown in Figure 7, our model demonstrates significant improvements over Standard ICL with various LLMs. Additionally, we observe that the performance of Standard ICL improves with the increasing size of LLMs, enhancing their reasoning capabilities. In contrast, our model achieves the superior performance of larger LLMs even when applied to smaller-sized ones.

B.3 Overhead Analysis

As our model is training-free, its computational overhead arises during the inference phase. We record the spatial (GPU usage of open-source LLMs) and temporal (Average iteration time per sample) overheads based on various LLMs: For Qwen2, the spatial overhead is 2.93 GB, 24.58 GB, and 76.50 GB for model sizes of 1.5B, 7B, and 72B, respectively, while the temporal overhead is 5.04 s, 8.07 s, and 77.67 s, respectively.

C Details of Datasets

The ECPE dataset classifies emotions into five categories: happiness, sadness, disgust, fear, and surprise. However, this classification lacks the necessary granularity to capture nuanced emotional states. For example, the emotion labeled as 'happiness' in the sentence 'somebody has earned everyone's respect' would be more precisely categorized as 'admiration'. To address this, we selected samples labeled as happiness, sadness, disgust, and fear from the dataset and manually mapped these categories to 12 emotions defined by the OCC model, such as joy, admiration, and gratitude[12]. To ensure fairness, we preserves the original annotations of both emotion and cause clauses. Specifically, Table 5 presents the mapping between emotion types in original dataset and emotion types in the OCC model, while Table 6 summarizes the sample distribution of these emotions in the ECPE and SECI dataset.

D Details of the Model Implementation

D.1 Data Prepocessing

Appraisal Perspective Extraction In ECE and ECPE tasks, the appraisal perspective is not predefined, making its extraction essential prior to reasoning. We employed the LLM-based Standard ICL method to extract the appraisal perspective, with the specific template prompts shown in Table 8. To maximize the accuracy of the reasoning results, this extraction process is repeated five times,

[5]For LLMs with size above 30B, we use 8 bit quantization.
[6]https://huggingface.co/Qwen/Qwen2-1.5B-Instruct
[7]https://huggingface.co/microsoft/Phi-3.5-mini-instruct
[8]https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct
[9]https://huggingface.co/01-ai/Yi-1.5-34B-Chat
[10]https://huggingface.co/Qwen/Qwen2-72B-Instruct
[11]https://platform.openai.com/docs/models/gpt-3-5-turbo

[12]Samples labeled as surprise were excluded due to ongoing debates surrounding the question, 'Is surprise an emotion?' Numerous studies argue that surprise is merely an intermediate cognitive state involved in environmental appraisal, rather than a fully-fledged emotion (Nomikos et al., 1968; Scherer, 1984; Lazarus, 1991).

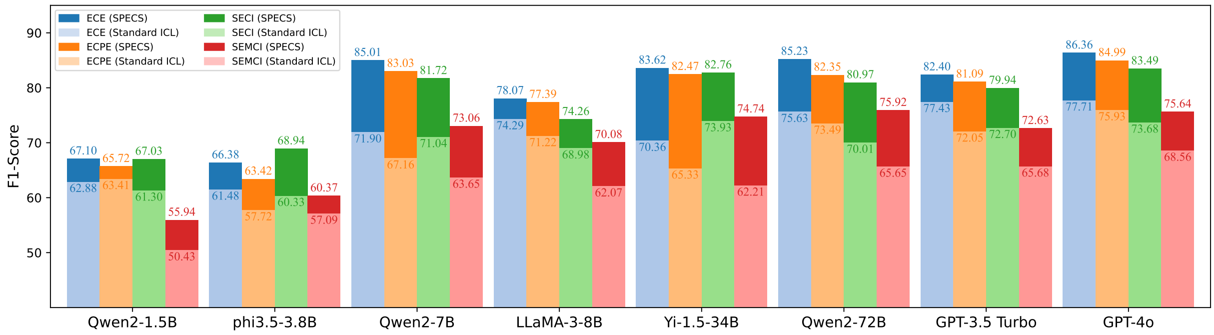

Figure 7: The performance of various LLMs across four emotion cause analysis tasks, with the performance of our SPECS model and Standard ICL being represented respectively.

and the response with the highest cumulative semantic similarity to all other responses is selected as the final answer.

Emotion Clauses Initialization Identifying emotion clauses is crucial in the ECPE task. Before applying explicit reasoning with LLM, we trained a simple discriminative model for emotion clause initialization. The model uses a BERT encoder to generate vector representations, followed by an MLP layer to identify emotion clauses. During hyperparameter tuning, we prioritized recall to ensure accurate identification of all emotion clauses. Misclassified non-emotion clauses are filtered during the Emotion Elicitation phase, improving precision. The ECPE dataset is split using 10-fold cross-validation, with one part as the test set and the remaining nine as the training set. The model achieved 82.37% precision and 97.74% recall in the initialization phase, with false positives later filtered by the *ECS* refinement process.

D.2 Hyper-parameters

Table 7 presents the hyper-parameter settings for the SPECS framework. The model's hyper-parameters include α: the threshold of cognitive consistency; δ_0: the weight of cognitive consistency in demonstrations selection; $\delta_1, \delta_2, \delta_3$: the weights of clause-level, text-level and scene-level context consistency in demonstrations selection; T_0: the number of iterations in *ECS* initialization phrase; T_1: the total number of iterations; $N_{Positive\ Demos}, N_{Negative\ Demos}$: the numbers of the positive and negative demonstrations; temperature: the temperature of LLMs.

D.3 Template prompts

Table 8 presents the template prompts for each subprocess in the SPECS framework during LLM-based reasoning, including the system instruction I for setting background information and task configuration, as well as the user question q for presenting the analysis task. The design of prompts is based on the rules from the OCC model (Ortony et al., 1990), including the definitions of appraisal dimension variables and their mappings to different emotion types. Building on this, for each reasoning subprocess in our model, we assessed from a group of candidate prompt templates and chose the best one as the final version.

Emotion Types in Original Dataset	Emotion-triggering Words	Emotion Types in the OCC model
Happiness	*Gao1 Xing4* (Happy), *Xi3 Yue4* (Joy), *Xing4 Fu2* (Happiness), ...	**Joy**
	Zun1 Jing4 (Respect), *Qin1 Pei4* (Admiration), *Zan4 Mei3* (Praise), ...	**Admiration**
	Jiao1 Ao4 (Pride), *Zi4 Hao2* (Proud), ...	**Pride**
	Gan3 Xie4 (Thank), *Gan3 Ji1* (Appreciate), *Gan3 En1* (Grateful), ...	**Gratitude**
	Man3 Zu2 (Gratification)	**Gratification**
Sadness	*Shang1 Xin1* (Sad), *Tong4 Ku3* (Suffering), *Yu4 Men4* (Depressed), ...	**Distress**
	Hou4 Hui3 (Regret), *Chan4 Hun3* (Repentance), *Ao4 Hui3* (Remorse), ...	**Remorse**
	Nei4 Jiu4 (Guilty), *Can2 Kui4* (Shame), *Zi4 Ze2* (Self-blame), ...	**Shame**
Disgust	*Fan2 Men4* (Anxiety), *Fan4 Chou2* (Worry), *Bu4 Gan1 Xin1* (Discontent), ...	**Distress**
	Bi3 Yi2 (Contempt), *Ze2 Bei4* (Blame), *Bu4 Man3* (Dissatisfaction), ...	**Reproach**
	Diu1 Lian3 (Lose face), *Xiu1 Kui4* (Shame), ...	**Shame**
	Fen4 Hen4 (Resentment), *Huai2 Hen4* (Bear a grudge), *Fan2 Zao4* (Irritation), ...	**Anger**
	Hui3 Hen4 (Remorse)	**Remorse**
Fear	*Hai4 Pa4* (Fear), *Kong3 Ju4* (Fright), *Jiao1 Lv4* (Anxiety), ...	**Fear**
	Da4 Ku1 (Sobbing), *Gan3 Jue2 Tian1 Yao4 Ta1* (Feel like the world is falling apart)	**Distress**
	Jiong3 Po4 (Awkwardness), *Xiu1 Kui4 Nan2 Dang1* (Overwhelming shame)	**Shame**
Anger	*Fen4 Nu4* (Anger), Qi4 Nao3 (Irritation), *Nu4 Huo3* (Fire of anger)	**Anger**
Surprise	*Cha4 Yi4* (Astonished), *Jing1 Ya4* (Surprise)	/

Table 5: Map the six emotion categories from the ECPE dataset to the twelve emotion categories in the OCC model. For each emotion mapping, several Emotion-triggering Words are provided as examples, along with the Chinese pinyin of the original words in the dataset and their corresponding English meanings.

Emotion	Sample Number	
	ECPE	SECI
Joy	497	250
Distress	565	250
Pride	9	0
Shame	64	0
Admiration	23	250
Reproach	17	250
Gratification	2	0
Remorse	78	0
Gratitude	19	250
Ange	318	250
Hope	0	0
fear	397	0

Table 6: A statistical summary of emotion samples in the ECPE and SECI datasets.

Hyper-parameters	Tasks			
	ECE	ECPE	SECI	SEMCI
α	0.9	0.9	0.85	0.85
δ_0	1	1	1	1
δ_1	1	1	1.2	1.2
δ_2	0.6	0.6	0.6	0.6
δ_3	2	2	1.2	1.2
T_0	5	5	5	5
T_1	10	10	10	10
$N_{Positive\ Demos}$	2	2	1	1
$N_{Negative\ Demos}$	1	1	2	2
temperature	1	1	1	1

Table 7: Hyper-parameters of our model

Operation	**Task**	**Prompt** $p = I\|\|\theta_1\|\|\theta_2\|\| \cdots \|\|q$	
		System Instruction I	**User Question** q
Appraisal Perspective Extraction	ECE ECPE	**System**: *"Your task is to extract entities from the text. Given a passage and an associated emotion, identify who in the text is expressing that emotion, and directly provide the person's name."*	**User**: *"Text:* [TEXT]*, who is experiencing the emotion described as* [Emotion]*? Please answer with the person's name directly."*
Cognitive Appraisal	ECE ECPE SECI SEMCI	**System**: *"You are a human with emotion cognitive capabilities. Given a passage of text, your task is to appraise the* [Desirability] / [Praise/Blame] / [Likelihood] *of a particular clause from the perspective of a specified individual."*	**User**: *"Text:* [TEXT], *decide whether clause* [c] *describes event/action* [S] *in a way that would be appraised as* [Desirable] / [Undesirable] / [Praise] / [Blame] / [Certain] / [Uncertain] *from* [p]*'s perspective."*
Emotion Elitation	ECPE SEMCI	**System**: *"You are a human with emotion cognitive capabilities. Given a passage of text, your task is to analyze whether a specific emotion is generated by the appraisal of a particular event or action from the perspective of a specified individual."*	**User**: *"According to OCC model, people feel* [Emotion] *in response to events/actions appraised as* [Appraisal]. *Given* [TEXT], *decide whether* [p] *may feel* [Emotion] *caused by* [C], *which describe a* [Appraisal] *event/action* [S]*."*
Cause Analysis	ECE ECPE SECI SEMCI	**System**: *"You are a human with emotion cognitive capabilities. Given a passage of text, your task is to analyze the reason for the emotion from the perspective of the specified person."*	**User**: *"According to OCC model, people feel* [Emotion] *in response to events/actions appraised as* [Appraisal]. *Given* [TEXT]*, in this passage,* [p] *appraises clauses like* [C] *as* [Appraisal]*, leading to the emotion of* [Emotion]. *Please summarize the specific event/action that caused this emotion."*

Table 8: Template prompts p for LLM-based entity extraction in the process of graph construction. D, E respectively denote the input slots for the Document and the Emotion. For the ECE task, the model extracts the emotion holder from a specified sentence, given the emotion E and its corresponding clause c_E. In the ECPE task, without predefined E and c_E, the model must identify the emotion holders for all emotions within the entire text.