MGCL: Multi-Granularity Clue Learning for Emotion-Cause Pair Extraction via Cross-Grained Knowledge Distillation

Yang Yu, Xin Lin*, Changqun Li, Shizhou Huang, Liang He
East China Normal University
{52205901014,52215901009,52275901004}@stu.ecnu.edu.cn
{xlin,lhe}@cs.ecnu.edu.cn

Abstract

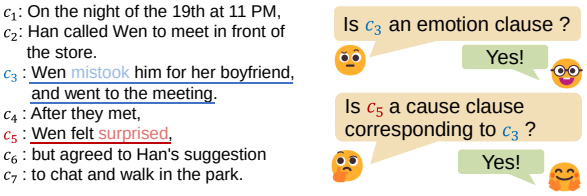
Emotion-cause pair extraction (ECPE) aims to identify emotion clauses and their corresponding cause clauses within a document. Traditional methods often rely on coarse-grained clause-level annotations, which can overlook valuable fine-grained clues. To address this issue, we propose **M**ulti-**G**ranularity **C**lue **L**earning (**MGCL**), a novel approach designed to capture fine-grained emotion-cause clues from a weakly-supervised perspective efficiently. In MGCL, a teacher model is leveraged to give sub-clause clues without needing fine-grained annotated labels and guides a student model to identify clause-level emotion-cause pairs. Furthermore, we explore domain-invariant extra-clause clues under the teacher model's advice to enhance the learning process. Experimental results on the benchmark dataset demonstrate that our method achieves state-of-the-art performance while offering improved interpretability.

1 Introduction

Investigating the underlying causes behind emotions is a promising research direction in sentiment analysis. The task of Emotion-Cause Pair Extraction (ECPE) (Xia and Ding, 2019) involves extracting all pairs of emotions and their corresponding causes within a document. Compared to the Emotion Cause Extraction (ECE) task first proposed by Lee et al. (2010), ECPE is significantly more challenging because the emotions in the document do not need to be pre-annotated. In other words, this task requires the simultaneous identification of all potential pairs, presenting a new challenge for document understanding capabilities.

Traditional ECPE methods can be categorized into two paradigms: two-stage methods and end-to-end methods. Xia and Ding (2019) proposed a two-stage framework involving the individual prediction of emotion and cause clauses, followed by filtering out incorrect pairs. However, this method introduces potential error propagation between stages. To address this issue, subsequent studies (Ding et al., 2020a,b; Wei et al., 2020) have investigated various end-to-end methods that directly determine the causal relationship between clauses. In contrast, the recent success of document-level machine reading comprehension (MRC) methods (Cheng et al., 2023; Zhou et al., 2022) suggests that using powerful pre-trained language models (PLMs) (Devlin et al., 2019) with appropriately fine-tuning sub-tasks can enable two-stage frameworks to achieve comparable performance. The significant performance gap observed between traditional two-stage methods and MRC-based methods remains unclear. We hypothesize that fine-tuning tasks at the clause level may further impact the performance of token-level encoders. Considering these factors, we adopt the MRC framework in this work.

Another significant research direction focuses on encoding clause-level features, including intra-



(a) (b)

Figure 1: An example: Figure 1a is an example document from the ECPE corpus. The emotion clause is underlined in red, and the cause clause is underlined in blue. Similarly, the words in red are the keywords about emotion, while the words in blue pertain to the cause. Figure 1b illustrates an example of an ECPE task decomposed into multiple turns using MRC. For ease of reading, we have translated the example from Chinese into English.

*Corresponding author.

Findings of the Association for Computational Linguistics: EMNLP 2024, pages 1897–1907

clause and inter-clause features. Recent works (Xia and Ding, 2019; Ding et al., 2020a,b; Cheng et al., 2023) utilize sequential encoders, such as Bi-directional LSTM, to encode task-specific clause features in narrative order. Following Wei et al. (2020); Chen et al. (2020), Graph Convolutional Network (GCN) (Kipf and Welling, 2017) and Graph Attention Network (GAT) (Veličković et al., 2018), have also been employed to capture features. In addition, leveraging auxiliary tasks like Emotion Clause Extraction (EE) and Cause Clause Extraction (CE) has been shown to enhance the transfer of information from emotion/cause encoders to their respective paired encoders. However, these encoders primarily focus on intra-clause features, while overlooking the inter-clause information. Liu et al. (2022) argues that this imbalanced information flow can lead to exposure issues among encoders. To address this, they constructed heterogeneous nodes with diversified edges (clause-clause, clause-pair, and document-clause-pair) for better inter-clause feature fusion. Despite these advancements, current research lacks studies exploring underlying fine-grained clues for identifying emotions, causes, and pairs.

We propose a multi-granularity clue learning (MGCL) method to address the aforementioned challenges. As illustrated in Figure 1, we further explore two types of fine-grained clues: sub-clause clues and extra-clause clues. Sub-clause clues are words or phrases directly related to the target task, such as emotion words in EE or action/event triggers in CE. While extra-clause clues, on the other hand, are domain-variant linguistic hints not directly related to the task's target and are often overlooked in coarse-grained annotations. Compared to inter-clause and intra-clause relationships, the boundaries of multi-granularity clues are more ambiguous and diverse. Intuitively, fully utilizing these clues can significantly aid the ECPE task. However, these fine-grained clues are not available in traditional ECPE settings.

In order to utilize fine-grained clues without explicit annotations, we propose a cross-granularity knowledge distillation method from a weakly-supervised learning perspective. Using coarse-grained annotations as supervision, an instance-level multiple instance learning (MIL) (Amores, 2013) framework is employed to learn fine-grained sub-clause clues. Importantly, this model can also directly explain the relevance between tokens and the target task. Even though explicit sub-clause labels remain ambiguous, inspired by knowledge distillation (KD) (Hinton et al., 2015), we use soft logits learned by the teacher model to guide the student model. Additionally, we use a clue-guided masking strategy to leverage extra-clause clues. Pseudo-examples are created by masking known sub-clause clues to guide the model with consistency alignment, even without explicit annotations. As a result, the model can learn to leverage these clues with only coarse-grained guidance and utilize sub-clause clues to help incorporate extra-clause features effectively.

Our contributions in this work can be summarized as follows:

- We introduce a novel multi-granularity clue learning (MGCL) to address the challenge of effectively capturing and leveraging multi-granularity clues in emotion-cause pair extraction.
- We develop a cross-grained knowledge distillation approach that enables the model to learn fine-grained clues effectively under the supervision of coarse-grained annotations.
- Experimental results on the benchmark dataset demonstrate significant improvements in ECPE performance by integrating sub-clause and extra-clause features. Additionally, fine-grained clues contribute to the interpretability of the model, offering more transparent insights into how emotions and their causes are identified within documents.

2 Related Work

Multiple Instance Learning Multiple instance learning (Amores, 2013) was widely applied in the field of weakly supervised learning, including applications such as fine-grained sentiment analysis (Angelidis and Lapata, 2018) and distantly supervised relation extraction (Mintz et al., 2009; Zeng et al., 2015). Unlike traditional supervised learning, where each instance in the training data is individually labeled, MIL organizes training data into "bags". Each bag contains a set of instances, and only a single label is assigned to each bag. In this work, the target sentences of any subtask are treated as bags, where we only have access to their corresponding coarse-grained annotations. The fine-grained sentence components, such as characters, words, or phrases, are treated as instances within these bags.

Knowledge Distillation Knowledge distillation (KD) was first proposed by Hinton et al. (2015) to improve the performance of a student model by leveraging the guidance of a teacher model. Building on this idea, Chen et al. (2024) implemented KD in the context of emotion-causal span pair extraction in conversations. In their approach, a teacher model predicted causal connective words between utterances, guiding the student model in identifying specific emotion labels and causal spans. To reduce the high computational costs of using a pre-trained teacher model, online knowledge distillation (Zhang et al., 2018) allows all participating models to learn from each other equally in a one-stage training process. Drawing inspiration from knowledge distillation and its variations, our method can learn directly from logits rather than explicit labels, thereby simplifying training and enhancing efficiency.

3 Problem Formulation

Given a document $D = (c_1, c_2, \ldots, c_{|D|})$ consisting of $|D|$ clauses, where i-th clause $c_i = (w_1^i, w_2^i, \ldots, w_{|c_i|}^i)$ contains $|c_i|$ words, the goal of ECPE is to extract all the emotion-cause pairs from the document D:

$$P = \{\ldots, (c_i^e, c_j^c), \ldots\} \quad (1 \leq i, j \leq |D|) \quad (1)$$

where c_i^e is c_i serving as a emotion clause, and c_j^c represents the corresponding cause clause in the pair.

Meanwhile, ECPE involves two auxiliary tasks: clause extraction (EE) and cause clause extraction (CE). A clause c_i is an emotion clause if any pair (c_i^e, c_j^c) is established. This can be formally defined as follows:

$$y_i^e = \begin{cases} 1, & \text{if } \left(c_i^e, c_j^c\right) \in P \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $y_i^e = 1$ indicates that c_i is an emotion clause. The extraction of cause clauses can be defined similarly:

$$y_j^c = \begin{cases} 1, & \text{if } \left(c_i^e, c_j^c\right) \in P \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

4 Methodology

4.1 Two-stage MRC Framework

The architecture of the MRC framework is illustrated in Figure 2. In the first stage, emotion or cause clauses from the document are detected in a question-answering fashion with task-specific pre-defined prompts. In the second stage, corresponding cause clauses or emotion clauses are further identified by emotion-specific or cause-specific prompts. Finally, diverse inference strategies, such as Rethink (Zhou et al., 2022) and Set Combination (Cheng et al., 2023), are used to give the final answer predictions.

Task Prompt	Template
$Prompt_{EE}$	This is an emotion clause
$Prompt_{CE}$	This is a clause emotion
$Prompt_{ESCE}$	The emotion clause c_i is corresponding to this cause clause
$Prompt_{CSEE}$	The cause clause c_i is corresponding to this emotion clause

Table 1: Task specific prompts for MRC.

Task Prompt Design Following previous works (Cheng et al., 2023; Zhou et al., 2022; Zheng et al., 2022), different types of prompts corresponding to the subtasks mentioned above are designed to formalize the ECPE task to the MRC task. As shown in Tab 1, four task prompts are listed as follows: $Prompt_{EE}$ is designed to extract all emotion clauses for Emotion Clause Extraction (EE) task; $Prompt_{CE}$ is designed to extract all cause clauses for Cause Clause Extraction (CE) task; $Prompt_{ESCE}$ is designed to extract all cause clauses corresponding to clause c_i for Emotion-Specific Cause Clause Extraction(CSEC) task; $Prompt_{CSEE}$ is designed to extract all emotion clauses corresponding to clause c_i for Cause-Specific Emotion Clause Extraction(CSEC) task.

Formally, given the task-specific pre-defined prompt $Prompt_*$ as query and linearized document $\mathbf{x}$ as context. The input sentence of the encoder can be denoted as:

$$\mathbf{x}_* = \text{[CLS]}\ \ Prompt_*\ \ \text{[SEP]}\ \ \mathbf{x} \quad (4)$$

where "[CLS]" and "[SEP]" are special tokens used in Devlin et al. (2019); $* \in \{EE, CE, ESCE, CSEE\}$; and $\mathbf{x}$ is a linearized text constructed by joining clauses in the document D by separation token:

$$\begin{aligned} \mathbf{x} = \{&x_1^q, x_2^q, \ldots, x_{|c_1|}^q, \text{[SEP]}, x_1^1, x_2^1, \\ &\ldots, \text{[SEP]}, x_1^{|D|}, x_2^{|D|}, \ldots, x_{|c_{|D|}|}^{|D|}\} \end{aligned} \quad (5)$$

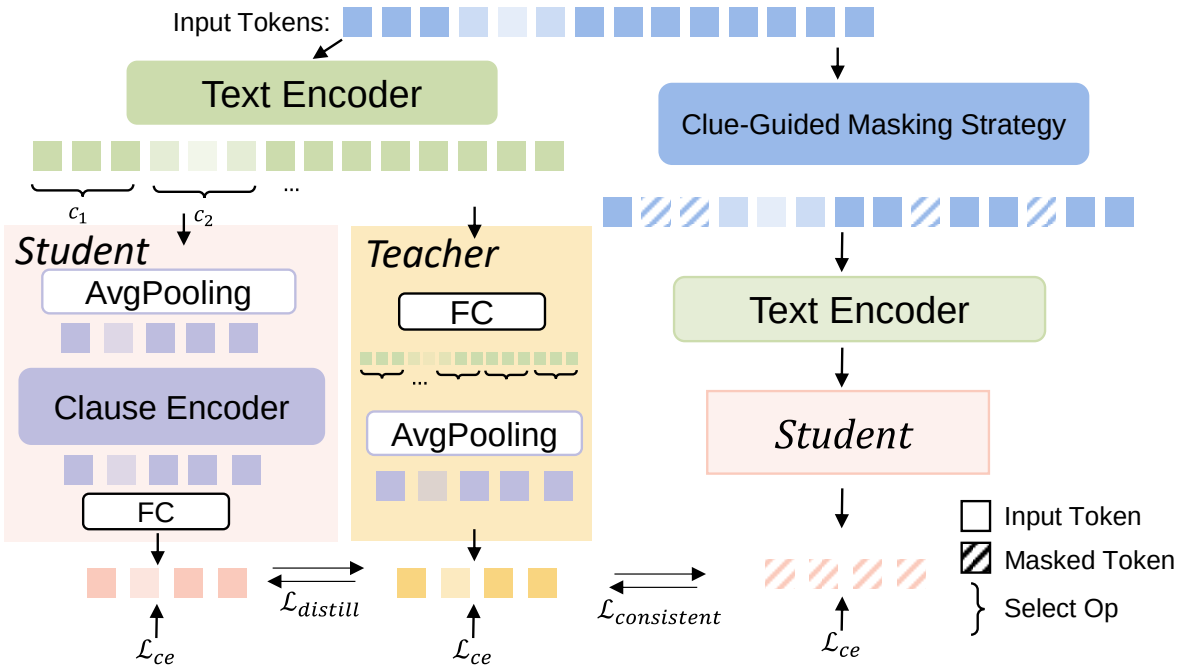


Figure 2: The overall architecture of the MGCL. Firstly, the text encoder generates representations of each token. These representations of tokens are then fed into the teacher and student models, which focus on coarse-grained and fine-grained clues respectively, to complete the ECPE task. Specifically, the representations are organized into corresponding clause groups by Select Op and then transformed into clause representations using AvgPooling at different stages. The models can learn from each other with KD. Simultaneously, the teacher model assigns importance scores to each token, which are used to mask the original input in the Clue-Guided Masking Strategy. Then the masked input is fed back into the student model for consistency comparison. Notably, all sub-models share learnable parameters.

where x_j^q is is the j-th token of the query; x_j^i is the j-th token of the i-th clause in the document D.

Document Encoder To capture such a "word-clause-document" structure, prior research has introduced various methods. In this work, we utilize a pre-trained BERT (Devlin et al., 2019) as the text encoder, which is adept at integrating representations at the word level:

$$\begin{aligned} H &= \text{TextEncoder}(D') \\ &= \{h_{[\text{CLS}]}, h_1^q, h_2^q, \ldots, h_{|q|}^q, h_{[\text{SEP}]}, h_1^1, \\ &\quad h_2^1, \ldots, h_{[\text{SEP}]}, h_1^{|D|}, h_2^{|D|}, \ldots, h_{|c_{|D|}|}^{|D|}\} \end{aligned} \tag{6}$$

where $H \in \mathbb{R}^{|D'| \times d}$ and d denotes the hidden size of the encoder; h_j^q is the hidden representation of the j-th token in the query; and h_j^i denotes the hidden representation of token x_j^i.

Then an average pooling is applied to the representations of tokens except for the [CLS] and [SEP] in each clause as the representations h_*^P of clauses:

$$\begin{aligned} H_P &= \text{AvgPooling}(H) \\ &= \{h_q^P, h_{c_1}^P, h_{c_2}^P, \ldots, h_{c_{|D|}}^P\} \end{aligned} \tag{7}$$

where $H_P \in \mathbb{R}^{(|D|+1) \times d}$ and h_q^P signifies the query's representation; And $h_{c_i}^P$ corresponds to the representation of the clause c_i.

Following previous works, a clause encoder is introduced to further fuse clause-level contextual information within the document. This can be achieved through various mechanisms such as Bi-LSTM (Xia and Ding, 2019; Ding et al., 2020a; Cheng et al., 2023), GCN (Chen et al., 2020; Wei et al., 2020; Liu et al., 2022), and its variants (Wei et al., 2020; Zhou et al., 2022). The document can be represented as follows:

$$\begin{aligned} H_C &= \text{ClauseEncoder}(H_P) \\ &= \{h_q^C, h_{c_1}^C, h_{c_2}^C, \ldots, h_{c_{|D|}}^C\} \end{aligned} \tag{8}$$

where $H_C \in \mathbb{R}^{(|D|+1) \times d}$. Specifically, h_q^C signifies the query's representation; And $h_{c_i}^C$ corresponds to the representation of the clause c_i.

Answer Prediction After obtaining the output representations of the last encoder layer, we concatenate the representation of their corresponding task-specific query and i-th clause together, and then feed them into a fully connected (FC) layer for the final prediction:

$$\hat{y}_i^* = W^T[h_q^C; h_{c_i}^C] + b \tag{9}$$

where $W \in \mathbb{R}^{2\times d}$ and $b \in \mathbb{R}^2$ are learnable parameters; $[\cdot;\cdot]$ is concatenation; and $\hat{y}_i^*$ denotes the logit for the task $*$.

Optimization The predicted probability p_i can be obtained by applying the softmax function to the logits $\hat{y}_i$. The cross-entropy loss for the task $*$ can be formulated as follows:

$$\mathcal{L}_{\text{CE}}^* = -\sum_{i=1}^{|D|} \hat{y}_i^* \log p_i \tag{10}$$

Inference For simplification purposes, the probability of each candidate task-specific answer can be defined as $P(c_i^e)$, $P(c_i^c)$, $P(c_j^c \mid c_i^e)$ and $P(c_j^e \mid c_i^c)$ for EE, CE, ESCE and CSEE respectively. If $P(\cdot) > 0.5$, the result is adopted. In this paper, we follow the rethink setting (Zhou et al., 2022), as it achieves the best results across all settings (Cheng et al., 2023; Zhou et al., 2022) in our experiments. In **Rethink** setting, the probability of the Emotion-Cause direction results ($P(c_i^e)P(c_j^c \mid c_i^e)$) is adjusted by a weight factor $\gamma(= 0.7)$ if they are not adopted by CSEE task. And the adjusted probability is given by $P(c_i^e, c_j^c) = \gamma P(c_i^e)P(c_j^c \mid c_i^e)$.

4.2 Multi-Granularity Clue Learning

To address the limitations of previous studies that relied solely on coarse-grained annotations as learning targets, we propose two additional types of fine-grained clues to improve learning.

Sub-Clause Clue This kind of clues provide fine-grained information about the role of a clause. For example, an emotion word can directly identify a clause as an emotion clause or a predicate can trigger an emotional response within a clause. Compared to clause-level annotations, the granularity of these clues is finer and can be in the form of tokens, phrases, or entire sub-clauses. However, for computational efficiency and practical implementation, we only focus on tokens in this paper.

Extra-Clause Clue In daily communication, we often don't need explicit evidence to identify the role of the clause. One way to do this is by looking at conjunctions and prepositions, which indicate the discourse information between clauses. Sometimes, clues may not appear directly within the target clause but can give hints about the role of the target clause from its surroundings. They can serve as supplementary components that suggest the presence of useful clues elsewhere in the text. For example, in the clause "Wen felt surprised", the word "felt" itself isn't an explicit emotional clue, but it indicates that emotional expression will come next. Therefore, this kind of clue is domain-invariant to emotion words.

We argue that these clues express clause relationships either directly or indirectly and can significantly assist in performing the ECPE task. However, previous works have not effectively utilized these fine-grained clues, partly due to the increased annotation cost associated with such detailed labeling. Therefore, exploring how to leverage these clues in the absence of explicit annotations remains a critical area of research. We will discuss this further in the next chapter.

4.3 Cross-Grained Knowledge Distillation

While we can't provide explicit fine-grained clues directly for supervision with clause-level annotations, inspired by MIL, we can obtain weakly supervised signals from coarse-grained annotations. By using knowledge distillation, these soft signals can be utilized to teach the model to leverage fine-grained clues effectively.

Teacher Model As illustrated in Figure 2, our approach differs from previous knowledge distillation methods in that it entails sharing all trainable parameters between the teacher and student models while utilizing different structures from each other. The teacher model treats each representation of each token as an instance and performs average pooling operations to aggregate the logit of all tokens within a clause:

$$\hat{y}_i^T = \frac{1}{|c_i|}\sum_{j}^{|c_i|} \hat{y}_{i,j}^T = \frac{1}{|c_i|}\sum_{j}^{|c_i|} W^T[h_q^C; h_j^i] + b \tag{11}$$

Based on our experience, this design helps students and teachers share general knowledge, while

maintaining differences across various granularities.

From Coarse to Fine Unlike mainstream KD approaches, explicit annotations for the aforementioned fine-grained clues are not available in our setting. Therefore, we employ a MIL approach to obtain pseudo-labels related to sub-clause clues in a weakly supervised manner. By distilling this soft-logit knowledge into the clause-level task, our method allows the model to effectively utilize these fine-grained clues, even when only coarse-grained labels are available. The process of transferring soft-logit knowledge at the instance level from the teacher to the student can be defined as follows:

$$\mathcal{L}_{\text{distill}}^{T\to S} = \mathcal{KL}(\sigma(\hat{y}_i^T)\|\sigma(\hat{y}_i)) \tag{12}$$

where $\mathcal{KL}(\cdot\|\cdot)$ stands for the KL divergence (Hinton et al., 2015); $\sigma(\cdot)$ is the softmax function with a temperature hyper-parameter t that softens the probability distributions:

$$\sigma(x_i) = \frac{\exp(x_i/t)}{\sum_j \exp(x_j/t)} \tag{13}$$

Inspired by Ji et al. (2021), the teacher can enhance itself through the diverse feedback from the student. Therefore, we transfer the knowledge from an additional direction $\mathcal{L}_{\text{distill}}^{S\to T}$ and the final distillation loss can be defined as follows:

$$\mathcal{L}_{\text{distill}} = 0.5 * (\mathcal{L}_{\text{distill}}^{T\to S} + \mathcal{L}_{\text{distill}}^{S\to T}) \tag{14}$$

Clue-guided Consistency Alignment Extra-clause clues provide indirect evidence for solving the task, allowing the determination of clause roles without relying on explicit clues. In other words, it is possible to determine the role without direct evidence, such as emotion words or predicate triggers. To improve the ability to utilize this kind of clues, we propose a clue-guided masking strategy to transform it into a consistency alignment task.

The text is firstly corrupted based on the guidance provided by the teacher model, with each corrupted token being replaced by a single "[MASK]" token. Specifically, the corrupted token is determined if teacher's confidence $\sigma(\hat{y}_{i,j}^T) > \alpha$ on the target token x_j^i. The student model then predicts the role of the corrupted clause, which should be consistent with the original text. To avoid overfitting on "[MASK]" tokens, we also apply a standard random masking strategy described in Devlin et al. (2019). Input tokens are randomly masked by a probability β. Formally, let $\mathbf{x}_{\text{corrupt}}$ be the corrupted text and $\hat{y}_i^C$ be the logits of the final prediction depending on $\mathbf{x}_{\text{corrupt}}$, we can define the consistent constraint as follows:

$$\begin{aligned}\mathcal{L}_{\text{consistent}} =& 0.5 * (\mathcal{KL}(\sigma(\hat{y}_i^C)\|\sigma(\hat{y}_i^T)) \\ &+ \mathcal{KL}(\sigma(\hat{y}_i^T)\|\sigma(\hat{y}_i^C)))\end{aligned} \tag{15}$$

In summary, the above steps incorporate multi-granularity clue learning to include sub-clause and extra-clause level clues, as well as cross-grained knowledge distillation to use these clues without explicit annotations effectively. This approach greatly improves the model's performance and interpretability in ECPE.

4.4 Training Objective

The overall loss of our proposed MGCL framework is given as:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{distill}} + \mathcal{L}_{\text{consistent}} \tag{16}$$

where $\mathcal{L}_{\text{CE}}$ represents the sum of cross-entropy losses for all tasks on different sub-models.

5 Experiments

5.1 Experimental Setup

Datasets In the paper, we conduct experiments on the benchmark dataset (Xia and Ding, 2019), which is constructed based on a public Chinese emotion corpus (Gui et al., 2016) from the SINA city news. The dataset contains 1,945 documents and 28,727 clauses. The number of candidate clause pairs is 490,367 and the number of valid emotion cause clause pairs is 2,167.

In the experiment, we use 10-fold cross-validation. Following the setting used by Xia and Ding (2019), we stochastically select 90% of the data for training and the remaining 10% for testing.

Metrics Following Xia and Ding (2019), we choose precision(P), recall(R), and F1-score(F1) as evaluation metrics across all tasks.

Baselines We compare our method with recent strong baselines, including: **ECPE-2D** (Ding et al., 2020a), **TransECPE** (Fan et al., 2020), **PairGCN** (Chen et al., 2020), **RANKCP** (Wei et al., 2020), **MM-R** (Zhou et al., 2022), **CD-MRC**(Cheng et al., 2023), **PBJE** (Liu et al., 2022), **JCB** (Feng et al., 2023). Among them, **MM-R** and **CD-MRC** convert ECPE to MRC task.

Model	ECPE			EE			CE		
	P	R	F1	P	R	F1	P	R	F1
ECPE-2D	72.92	65.44	68.89	86.27	92.21	89.10	73.36	69.34	71.23
TransECPE	77.08	65.32	70.72	88.79	83.15	85.88	78.74	66.89	72.33
PairGCN	76.92	67.91	72.02	88.57	79.58	83.75	79.07	68.28	73.75
RANKCP†	71.19	76.30	73.60	91.23	89.99	90.57	74.61	77.88	76.15
PBJE	79.22	73.84	76.37	90.77	86.91	88.76	81.79	76.09	78.78
JCB	79.10	75.84	77.37	90.77	87.91	89.30	81.41	77.47	79.34
MM-R†	82.18	79.27	80.62	97.38	90.38	93.70	83.28	79.64	81.35
CD-MRC†	82.49	78.00	80.13	96.92	**93.98**	**95.37**	81.01	**80.68**	80.77
MGCL	75.83	79.09	77.36	88.49	90.94	89.66	78.57	80.64	79.53
MGCL†	**83.41**	**80.13**	**81.66**	**97.94**	91.57	94.62	**84.26**	80.61	**82.32**

Table 2: Experimental results of on ECPE benchmarks. The best result is marked in **bold** and the second-best performance is underlined. † indicates that the model employs an emotion-filtering strategy with a sentiment lexicon.

Implementation Details All implementations in this paper are built on the top of PyTorch (Ansel et al., 2024), Transformers (Wolf et al., 2020), and PyG (Fey and Lenssen, 2019) on GeForce RTX 3090 GPUs. During training, we employed the AdamW optimizer and a linear learning rate scheduler with a warm-up setting. We set the batch size to 8 and the max epochs to 10 epochs. In addition, reported results are medians over 5 random initializations (seeds) for a fair comparison. Following comparable baselines, the pre-trained BERT is initialized with checkpoint "bert-base-Chinese"[1]. Following previous works (Wei et al., 2020; Cheng et al., 2023; Zhou et al., 2022), ANTUSD (Wang and Ku, 2016) is used as the sentiment lexicon to determine whether the clause contains any sentiment word. The hyper-parameters specified in section 4.3 are assigned the values $\alpha = 0.5$, $\beta = 0.15$, and $t = 5$.

5.2 Main Result

Table 2 presents the experimental results across three tasks: ECPE, EE, and CE. The overall results demonstrate the effectiveness of the proposed MGCL compared to other baselines. MGCL shows a significant improvement in the ECPE task due to its ability to learn and utilize multi-granularity clues. Although methods employing sentiment lexicons achieve higher performance in the EE task due to incorporating external knowledge, they do not maintain the same advantage in the CE task. MGCL not only achieves the best performance on the primary ECPE task but also shows significant improvement in the CE task. This superior performance is attributed to the model's capability to efficiently learn and leverage multi-granularity clues. Importantly, this advantage is even more pronounced for methods that do not use sentiment lexicons. Owing to its ability to identify more potential target clauses, MGCL significantly improves recall across different tasks.

Compared to the most advanced method JCB, our method has a slightly lower F1 score on ECPE task. We argue that the frustration can be attributed to the optional relative distance constraint setting mentioned in Feng et al. (2023). It is worth noting that our method is position-insensitive (Bao et al., 2022), whereas Feng et al. (2023) has demonstrated that JCB's performance significantly drops (77.31 $\rightarrow$ 75.09 on F1) without the constraint. Similar observations have also been observed on RankCP and PBJE.

5.3 Ablation Study

Ablation studies are conducted to verify the effectiveness of MGCL. The results of the ablation study are shown in Table 3.

w/o Sub-Clause Clue We block knowledge transfer between the teacher and the student in this setting. Relying solely on clue-guided consistency alignment ($\mathcal{L}_{\text{consistent}}$), the performance is even weaker than in the w/o Cross-Grained KD setting. We argue that this phenomenon is related to the diminished ability to utilize sub-clause clues in the current setting. In other words, poor clue detection

[1]https://huggingface.co/google-bert/bert-base-chinese

Model	ECPE			EE			CE		
	P	R	F1	P	R	F1	P	R	F1
MGCL	75.83	**79.09**	**77.36**	88.49	**90.94**	**89.66**	78.57	**80.64**	**79.53**
w/o Sub-Clause Clue	75.90	78.54	77.10	88.58	89.64	89.04	78.75	80.19	79.36
w/o Extra-Clause Clue	75.98	78.84	77.30	89.07	89.81	89.40	78.82	80.24	79.46
w/o Cross-Grained KD	**76.00**	78.25	77.03	**89.49**	89.58	89.50	**78.89**	79.75	79.25

Table 3: The results of the ablation study on the benchmark for the main task and auxiliary tasks.

Model	ECPE		
	P	R	F1
MGCL(BERT Only)	74.51	**77.89**	**76.06**
w/o Sub-Clause Clue	75.60	76.23	75.80
w/o Extra-Clause Clue	**76.25**	75.46	75.79
w/o Cross-Grained KD	75.52	76.03	75.68
BERT	74.65	77.42	75.94
BERT + GCN	73.35	79.87	76.36

Table 4: The results without clause encoder for ECPE.

Method	Document
MIL	...c_2:Han called Wen to meet in front of the store. c_3:Wen mistook him for her boyfriend, and went to the meeting. c_4: After they met, c_5: Wen felt surprised...
MGCL	...c_2:Han called Wen to meet in front of the store. c_3:Wen mistook him for her boyfriend, and went to the meeting. c_4: After they met, c_5: Wen felt surprised...

Table 5: Visualization of task-specific token probability in MIL and MGCL. Red words indicate emotion, blue words indicate cause. Deeper colors represent higher probabilities.

leads to nearly random or even worse clue detection results.

w/o Extra-Clause Clue In this setting, only sub-clause clues ($\mathcal{L}_{\text{distill}}$) are utilized. The omission of extra-clause clues leads to an expected drop in performance, highlighting the importance of incorporating extra-clause clues for a more comprehensive understanding of the text.

w/o Cross-Grained KD This setting removes all cross-grained interaction ($\mathcal{L}_{\text{distill}}$ and $\mathcal{L}_{\text{consistent}}$), resulting in the framework degrading into multi-task learning with MIL. Without multi-granularity clues, the performance of Pair Extraction dramatically drops.

In summary, the ablation study validates the effectiveness of the multi-granularity clues and cross-grained knowledge distillation strategies employed in the MGCL framework, confirming their importance in enhancing the performance of ECPE tasks.

5.4 Effective Encoder Learning

To verify the effect of taking into account fine-grained clues in ECPE, we further conduct some experiments in special settings. As shown in Table 4, in the two most basic settings, **BERT** and **BERT + GCN**, the performance achieved is quite similar. In previous works, although the clause encoder has almost become a standard component, the performance improvement it provides has not been as significant as expected. We attribute this to the insufficient training of the clause encoder. By comparing different settings, we find that the main performance boost of MGCL stems from improving the task learning efficiency of the clause encoder. As illustrated in Table 4, MGCL has little impact on models without a clause encoder but significantly improves the performance of models with a clause encoder. The empirical evidence demonstrates that MGCL greatly enhances the learning capability of the clause encoder.

5.5 Case Study

We analyze an example selected from the benchmark corpus to demonstrate the effectiveness of MGCL, which is shown in Table 5. In the example, this document has a ground-truth emotion-cause pair (c_5, c_3). Compared to the MIL method, MGCL offers clearer boundaries for the clues, indicating that the model is more confident in its judgments and assigns lower probabilities to noise. Addition-

ally, the clues identified by our method exhibit better continuity and maintain semantic integrity. This suggests that multi-granularity clue learning can benefit the model in the ability to identify emotion and cause. Unlike previous work, our approach is the first to provide direct interpretability for predictions by discovering fine-grained clues.

6 Conclusion

In this paper, we proposed a **M**ulti-**G**ranularity **C**lue **L**earning (**MGCL**) method, which efficiently captures multi-granularity emotion-cause clues. From a weakly-supervised perspective, we introduce cross-grained knowledge distillation to learn fine-grained knowledge from coarse-grained annotations. The remarkable performance on the benchmark corpus demonstrates fine-grained clues can significantly assist in performing the ECPE task.

Limitations

Despite our progress, our work still has some limitations:

Firstly, following previous works, we implement MGCL within an MRC setting, which requires at least two rounds to complete the whole inference. Each inference is relatively independent, what means that even when performing the ECPE task on the same document, these predictions remain unaware of each other. This lack of global consideration during decision-making can result in suboptimal outcomes.

Secondly, data imbalance still presents a significant challenge to current works. Correct pairings may constitute less than 1% of the total data, and the imbalance is particularly problematic for MRC-based methods. The imbalance exponentially increases the number of available training samples, causing valuable positive samples to be overwhelmed by a vast number of negative samples.

Addressing these limitations is crucial for future exploration. Developing methods that can consider global context and dependencies between predictions could improve the overall performance. Additionally, strategies to handle data imbalance more effectively, such as advanced sampling techniques or data augmentation, could enhance the robustness and accuracy of the model.

Ethics Statement

This work complies with the ethics of ACL. The scientific artifacts we used are available for research with permissive licenses. The use of the artifacts in this paper adheres to their intended use and we do not believe the work presented here further amplifies biases already present in the datasets. Therefore, we foresee no ethical concerns in this work.

Acknowledgements

This work is supported by National Science and Technology Major Project (2021ZD0111000/2021ZD0111004), the Science and Technology Commission of Shanghai Municipality Grant (No. 21511100101, 22511105901, 22DZ2229004), the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education. Xin Lin is the corresponding author. Xin Lin is also a member of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education.

References

Jaume Amores. 2013. Multiple instance classification: Review, taxonomy and comparative study. *Artif. Intell.*, 201:81–105.

Stefanos Angelidis and Mirella Lapata. 2018. Multiple instance learning networks for fine-grained sentiment analysis. *Transactions of the Association for Computational Linguistics*, 6.

Jason Ansel, Edward Yang, Horace He, Natalia Gimelshein, Animesh Jain, Michael Voznesensky, Bin Bao, Peter Bell, David Berard, Evgeni Burovski, Geeta Chauhan, Anjali Chourdia, Will Constable, Alban Desmaison, Zachary DeVito, Elias Ellison, Will Feng, Jiong Gong, Michael Gschwind, Brian Hirsh, Sherlock Huang, Kshiteej Kalambarkar, Laurent Kirsch, Michael Lazos, Mario Lezcano, Yanbo Liang, Jason Liang, Yinghai Lu, CK Luk, Bert Maher, Yunjie Pan, Christian Puhrsch, Matthias Reso, Mark Saroufim, Marcos Yukio Siraichi, Helen Suk, Michael Suo, Phil Tillet, Eikan Wang, Xiaodong Wang, William Wen, Shunting Zhang, Xu Zhao, Keren Zhou, Richard Zou, Ajit Mathews, Gregory Chanan, Peng Wu, and Soumith Chintala. 2024. Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS 24)*. ACM.

Yinan Bao, Qianwen Ma, Lingwei Wei, Wei Zhou, and Songlin Hu. 2022. Multi-granularity semantic aware

graph model for reducing position bias in emotion cause pair extraction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1203–1213, Dublin, Ireland. Association for Computational Linguistics.

Xinhao Chen, Chong Yang, Changzhi Sun, Man Lan, and Aimin Zhou. 2024. From coarse to fine: A distillation method for fine-grained emotion-causal span pair extraction in conversation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(16):17790–17798.

Ying Chen, Wenjun Hou, Shoushan Li, Caicong Wu, and Xiaoqiang Zhang. 2020. End-to-end emotion-cause pair extraction with graph convolutional network. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 198–207, Barcelona, Spain (Online). International Committee on Computational Linguistics.

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 Trans. Inf. Syst.*, 41(4).

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.

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.

Chuang Fan, Chaofa Yuan, Jiachen Du, Lin Gui, Min Yang, and Ruifeng Xu. 2020. Transition-based directed graph construction for emotion-cause pair extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3707–3717, Online. Association for Computational Linguistics.

Huawen Feng, Junlong Liu, Junhao Zheng, Haibin Chen, Xichen Shang, and Qianli Ma. 2023. Joint constrained learning with boundary-adjusting for emotion-cause pair extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1118–1131, Toronto, Canada. Association for Computational Linguistics.

Matthias Fey and Jan E. Lenssen. 2019. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016. Event-driven emotion cause extraction with corpus construction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1639–1649, Austin, Texas. Association for Computational Linguistics.

Geoffrey E. Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. *CoRR*, abs/1503.02531.

Mingi Ji, Seungjae Shin, Seunghyun Hwang, Gibeom Park, and Il-Chul Moon. 2021. Refine myself by teaching myself: Feature refinement via self-knowledge distillation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10659–10668.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*.

Sophia Yat Mei Lee, Ying Chen, and Chu-Ren Huang. 2010. A text-driven rule-based system for emotion cause detection. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 45–53, Los Angeles, CA. Association for Computational Linguistics.

Junlong Liu, Xichen Shang, and Qianli Ma. 2022. Pair-based joint encoding with relational graph convolutional networks for emotion-cause pair extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5339–5351, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011, Suntec, Singapore. Association for Computational Linguistics.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *International Conference on Learning Representations*.

Shih-Ming Wang and Lun-Wei Ku. 2016. ANTUSD: A large Chinese sentiment dictionary. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2697–2702, Portorož, Slovenia. European Language Resources Association (ELRA).

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, Online. Association for Computational Linguistics.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. 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 Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 1003–1012. Association for Computational Linguistics.

Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1753–1762, Lisbon, Portugal. Association for Computational Linguistics.

Ying Zhang, Tao Xiang, Timothy M. Hospedales, and Huchuan Lu. 2018. Deep mutual learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

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.

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, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.