Causal Narrative Comprehension: A New Perspective for Emotion Cause Extraction

Wei Cao, Kun Zhang, Shulan Ruan, Hanqing Tao, Sirui Zhao, Hao Wang, Qi Liu, *Member, IEEE*, and Enhong Chen, *Senior Member, IEEE*

Abstract—Emotion Cause Extraction (ECE) aims to reveal the cause clauses behind a given emotion expressed in a text, which has become an emerging topic in broad research communities, such as affective computing and natural language processing. Despite the fact that current methods about the ECE task have made great progress in text semantic understanding from lexicon- and sentence-level, they always ignore the certain causal narratives of emotion text. Significantly, these causal narratives are presented in the form of semantic structure and highly helpful for structure-level emotion cause understanding. Nevertheless, causal narrative is just an abstract narratological concept and its involving semantics is quite different from the common sequential information. Thus, how to properly model and utilize such particular narrative information to boost the ECE performance still remains an unresolved challenge. To this end, in this paper, we propose a novel Causal Narrative Comprehension Model (CNCM) for emotion cause extraction, which learns and leverages causal narrative information smartly to address the above problem. Specifically, we develop a Narrative-aware Causal Association (NCA) unit, which mines the narrative cue about emotional results and uses the semantic correlation between causes and results to model causal narratives of documents. Besides, we design a Result-aware Emotion Attention (REA) unit to make full use of the known result of causal narrative for multiple understanding about emotional causal associations. Through the ingenious combination and collaborative utilization of these two units, we could better identify the emotion cause in the text with causal narrative comprehension. Extensive experiments on the public English and Chinese benchmark datasets of ECE task have validated the effectiveness of CNCM with significant margin by comparing with the state-of-the-art baselines, which demonstrates the potential of narrative information in long text understanding.

Index Terms—Emotion cause extraction, causal narrative, attention mechanism, semantics understanding

1 Introduction

As a sub-task of emotion analysis [1], Emotion Cause Extraction (ECE) has aroused extensive research interests in recent years [2], [3], [4]. It has significant potentials in various research communities (*e.g.*, affective computing [5] and natural language processing [6]) and wide applications (*e.g.*, social media and business intelligence [7], [8]). The key goal of ECE lies in identifying the cause clauses from a text with a given emotion. Fig. 1 illustrates an example of the ECE task in this paper.

Traditional approaches to the ECE task include rule-based methods [9], [10] and machine learning methods [11], [12]. Through making full utilization of linguistic rules and feature engineering, these methods have achieved quite good results in earlier years. However, most of these traditional methods still have limitations due to the lack of semantic understanding of the emotional text. With the development of deep learning technologies, many deep neural models [13], [14] have been introduced into ECE, attempting to alleviate the above problem by performing an in-depth semantic understanding of emotional context. The representative studies such as RTHN [14], MANN [15] which integrated attention mechanism into Recursive Neural Networks (RNN) [16] or Convolutional Neural Network (CNN) [17] for better text representation and emotion cause detection. Besides, there are also some other valuable attempts, such as the causes boundaries detection model SECA [18], knowledge fusion method RHNN [19] and so on. With these efforts and contributions, ECE has been pushed a large step forward.

However, most current methods pour attention to text semantic understanding from word-level and sentence-level, while ignoring causal narrative comprehension of the causal texts. Actually, each causal text for ECE usually contains multiple clauses (*i.e.*, emotion result, emotion cause, and others) that naturally have complex semantics as well as certain causal narrative structures. Moreover, these

- *Wei Cao is with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China, and also with Xinjiang Normal University, Urumqi, Xinjiang 830054, China. E-mail: cw0808@mail.ustc.edu.cn.*
- *Kun Zhang is with the Hefei University of Technology, Hefei, Anhui 230002, China. E-mail: zhang1028kun@gmail.com.*
- *Shulan Ruan, Hanqing Tao, Sirui Zhao, Hao Wang, Qi Liu, and Enhong Chen are with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China. E-mail: {slruan, hqtao, sirui}@mail.ustc.edu.cn, {wanghao3, qiliuql, cheneh}@ustc.edu.cn.*

Manuscript received 23 January 2022; revised 8 August 2022; accepted 4 September 2022. Date of publication 15 September 2022; date of current version 15 November 2022.
This work was supported in part by the National Natural Science Foundation of China under Grant 61727809, in part by the Young Scientists Fund of the National Natural Science Foundation of China under Grant 62006066, in part by the Open Project Program of the National Laboratory of Pattern Recognition (NLPR), and in part by the Fundamental Research Funds for the Central Universities under Grant JZ2021HGTB0075.
(Corresponding author: Enhong Chen.)
Recommended for acceptance by E. Cambria.
Digital Object Identifier no. 10.1109/TAFFC.2022.3206960

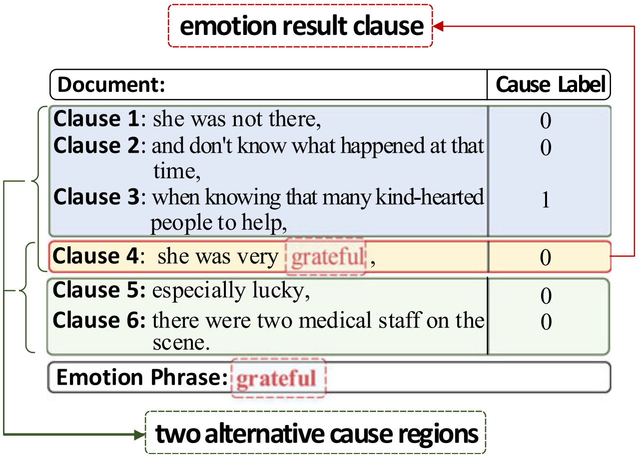


Fig. 1. An instance of the emotion cause extraction task. The document contains six clauses including one emotion phrase (*i.e.*, "*grateful*") indicating the overall emotion. Thus, we regard *clause 4* as result clause and aim to find out the cause clause (*i.e.*, *clause 3*) among all the clauses in this text.

causal narratives could have a considerable impact on text comprehension[20], [21]. In literature, causal narrative refers to the statement about causality during event evolution [22]. It influences the way of humans to conceptualize events [23], [24], [25] and contributes to the representation and comprehension of long texts [26], [27]. Generally, causal narratives could be divided into chronological narrative (cause–before–effect) and flashback narrative (effect–before–cause) according to narratology [22], [28]. Taking Fig. 1 as an example, according to the cue of the given emotion phrase *"grateful"*, *clause 4* could be regarded as the emotion result clause in the document. Thus, we could preliminarily locate the possible area of emotion causes, namely two alternative cause regions (*i.e.*, *clause 1-4* and *clause 4-6*). Further, with the causal narrative, we could efficiently mine emotional causal correlations between the result clause and other clauses for the ECE task.

Furthermore, in literature, some studies have also observed that there is an inherently strong correlation and coherence between the cause and result in a causal narrative [29], [30], [31], [32]. They demonstrate the result clause is most associated with the unrecognized cause clauses within the alternative cause regions. Based on this assertion, the grasp of causal narratives and semantic relations between result clauses and other clauses within causal narratives are considerable critical for causal text understanding [21], [33]. Consequently, we in this paper focus on causal narrative comprehension and exploring the emotional semantics correlations within causal narratives for better emotion cause extraction.

Inspired by the above observations, the specific solution in this paper includes two aspects. For one thing, we leverage the causal structure of causal narrative to perceive the possible scope of emotion cause clauses. For another, based on the guidance of causal structure, we focus on the clauses that have strong causal correlations with the known emotion result clause in a causal narrative to predict emotion cause clauses. To achieve the above solutions, we must consider the following challenges: 1) How to properly represent the textual causal structure via the causal narrative understanding of a document; 2) Under the guidance of causal narrative, how to explore and understand the causal association between cause clauses and result clauses within the document for emotion cause extraction.

To address the above challenges, in this paper, we propose a Causal Narrative Comprehension Model (CNCM) for emotion cause extraction. For the first challenge, we design a Narrative-aware Causal Association (NCA) unit, which uses the narrative cue about the known emotion result to learn the semantic correlation between causes and results for causal narrative representation. For the second challenge, we develop a Result-aware Emotional Attention (REA) unit to acquire the cognition of emotional causal correlation through the attention mechanism between the known result clause and other clauses within the causal narrative. Specifically, the REA unit is firstly performed for the preliminary cognition of emotional causal correlation of documents. With this preliminary cognition, we utilize the NCA unit for the representation of causal narrative structure for good comprehension of causality and perception about the possible scope of cause clauses. Third, guided by the representation of the causal narrative structure, the REA unit is performed again to acquire accurate comprehension of the emotional causal correlation for the prediction of emotion cause clauses.

As an emphasis, the main contributions of our work can be concluded as follows:

- We propose a model based on causal narrative comprehension for emotion cause extraction. To the best of our knowledge, it is the first time to introduce causal narrative information into the ECE task.
- We develop NCA to analyze and model the causal narrative information of ECE documents. Then we utilize REA to help understand the emotional causal correlations guided by the causal narrative information. In this way, we can grasp causality and identify the emotion causes of documents accurately.
- The experimental analysis of results on the benchmark datasets validates the effectiveness of the proposed CNCM for ECE. And the model achieves considerable performance by comparing with several state-of-the-art methods.

The remainder of this paper is organized as follows. The related work is introduced in Section 2, and the problem definition is stated in Section 3. Then, in Section 4, we demonstrate the model details and training techniques of the developed CNCM. Subsequently, the experimental results are reported in Section 5. Finally, we conclude this paper in Section 6.

2 Related Work

In this section, we will review the related works from three aspects: *Emotion Analysis*, *Emotion Cause Extraction* and *Narrative Understanding*, which are closely related to our work in this paper.

2.1 Emotion Analysis

With the boom of the artificial intelligence field, emotion analysis has attracted a large of research attention. In the field of text emotion analysis, scholars tend to mine effective text semantic information to improve the accuracy of emotion recognition [34], [35], [36]. For example, Li et al. [37] focused on the strong context dependence of each sentence in a discourse. They designed an appropriate framework named bidirectional emotional re-current unit (BiERU) to effectively encode the strong contextual information for conversational sentiment analysis. Simultaneously, there are also many meaningful explorations on visual emotion analysis [1], [38], [39]. For example, Ruan et al. [5] proposed a novel architecture named color enhanced cross correlation net (CECCN) for image sentiment analysis. As multimedia technology advances, multimodal data is rapidly growing and available to scholars. Many researchers began to pay attention to the research of multimodal emotion analysis [40], [41], [42]. Zhang et al. [43] noted that emotion in conversation videos happens step by step. Thus, they proposed a multimodal emotion recognition model based on reinforcement learning and domain knowledge for conversation videos. All these efforts exploit the characteristics of emotion information from various meaningful perspectives to promote emotion analysis and task implementation in certain scenarios.

2.2 Emotion Cause Extraction

As an important sub-task of emotion analysis, there has been an increasing amount of literature on ECE in recent years. Generally, previous methods can be grouped into three categories, *i.e., rule-based methods, machine learning methods* and *deep learning methods*.

2.2.1 Rule-Based Methods

Earlier studies of ECE are mainly rule-based methods [9], [11], [44]. Lee et al. [44] pioneered the construction of a publicly available corpus for the ECE task and conducted a detailed analysis of its content. To make full utilization of the special language expressions in this corpus to detect emotional causes, Lee et al. [11] generalized sets of linguistic rules well by defining linguistic cues. Subsequently, they further extracted cause expressions and specific constructions via linguistic rules to improve their previous solution [9]. In addition, there are also some novel solutions based on events analysis [45] and common sense knowledge [46], which have also been demonstrated quite good performance for ECE. However, these methods are not sufficiently generalized for practical applications, since artificial rules cannot cover all complex linguistic phenomena of texts in real situations.

2.2.2 Machine Learning Methods

Considering the weak generalization capability of rule-based methods, a variety of machine learning methods have also been developed for this task successively. For instance, to deal with the special linguistics features of Weibo [1] texts, Gao et al. [10] proposed a conditional random fields model based on syntactic and semantic characteristics, which could effectively mine the relation between emotion expression and cause in Weibo text to detect emotion causes of social texts. Then, Gao et al. [47] presented an Emotion-Cause-OCC model to address emotion cause extraction in micro-blog posts. Specially, this approach focused on investigating factors for eliciting kinds of emotions and could acquire the proportions of these cause components under different emotions. Additionally, there have also been many other valuable studies, such as the CRF-based model [12], the event-driven multi-kernel SVMs method [48]. All these studies have made large contributions to the development of ECE. However, these methods rely on the utilization of effective statistical features about the texts, ignoring text semantics understanding.

2.2.3 Deep Learning Methods

Owing to the development of deep learning technology, deep neural networks have attracted more and more research attention for their excellent semantic representation ability. Particularly, the Bi-directional Long Short-Term Memory (BiLSTM) [49] and attention mechanism are widely used in these studies since they could model good semantics and capture useful emotional information for the ECE task. For example, based on a co-attention deep neural network, Li et al. [50] took account into attention mechanism [51] and proposed a co-attention deep neural network to exploit the correlation among clauses which is helpful for emotion cause extraction. Following this work, MANN was proposed in [15], which substituted multi-attention-based framework for a co-attention network to mine correlations between emotion phrases and candidate clauses and achieved comparable results. To further improve the performance of the ECE task, Fan et al. [19] presented a novel solution called RHNN, which ingeniously utilized sentiment lexicon and common knowledge as restrained parameters to promote model training. Owing to the superiority of deep semantic representations and attention mechanisms, these models have gained great performance improvement. In addition, some other novel solutions have also been proposed, including hierarchical network methods [13], [14], question-answering solution [52], reordered prediction framework [53], retrieval rank framework [2] and document-level context idea [54], which have also provided some new insights for this task. Notably, there are also some derivative tasks about the ECE task. For example, some researchers innovatively improved the benchmark corpus of the ECE task to accommodate the derivative task of emotion-cause pair extraction and acquired many valuable results [55], [56], [57], [58], [59]. While some others addressed the ECE task as a boundary detecting task of emotion cause spans at the span-level by manually annotating cause spans on the original datasets [18].

To summarize, current advanced works about the ECE task have emphasized semantic representations of sentences [15], [19], [60]. They employed attention mechanisms to obtain emotion correlations based on emotion phrases for emotion cause extraction while ignoring causal narratives of documents. Unlike the above studies, in this paper, we deal with this task as an issue of causal narrative comprehension for documents. Particularly, we dig deeply into documents information to subtly model causal narratives of documents. In this way, we can efficiently localize emotion

1. https://weibo.com/

TABLE 1
Mathematical Notations

Symbol	Description
D	the original document with several clauses
ep	the emotion phrase of D
c_k^e	the k^{th} clause of D that contains emotion phrase ep, namely the emotion result clause of D
y	the cause label of the clauses in D
α	the emotional attention vector in REA
r	the causal narrative association vector of D
E_b	the feature representation of D with BERT
E_h	the hidden state by processing $\boldsymbol{E}_b$ with LSTM in REA
E_u	the representation of D that contains the preliminary Emotion Causality Understanding by REA
E_c	the representation of D that fuses the representation $\boldsymbol{E}_u$ and the causal narrative information $\boldsymbol{r}$
E_{ru}	the final representation of D with Emotion Causality Re-understanding

cause regions and accurately acquire emotional causal correlations, facilitating the task of emotion cause extraction.

2.3 Narrative Understanding

Narrative understanding aims to identify the key elements of narrative structure in a story, that is, the relationship between critical elements and context to which the narrative belongs [61]. It is usually applied to some understanding tasks, such as film analysis [62], [63], text extraction [64], [65], [66], reading comprehension [67], [68], [69], and so on. Generally, the narrative content of a text involves descriptions of daily activities, discourses, or stories [70]. While given the development logic of events and the completeness of discourses, the sentences within a narrative text (*e.g.*, a discourse or story) must show coherence in semantics [30], [32]. Nowadays, coherence has been introduced in many tasks related to narrative understanding and promotes these studies [71], [72], [73]. Similar studies have also been done in other fields such as the research about coherence in music generation [74] and the work about viewpoint coherence in film [75]. Considering that coherent narrative is bound to be semantically relevant, some tasks tend to deal with narrative coherence from the perspective of semantic correlation. For example, Hu et al. [76] proposed a model based on manual rules, which utilized causal potential to conclude event pairs with narrative causality relations from film scenes. Further, Chen et al. [73] focused on modeling sequential semantics of clauses in documents for story completing. Notably, as a special narrative structure, the causal narrative also presents a strong semantic coherence in its sentence sequence. As mentioned by Wellner et al. [31], the cause and result within a causal narrative are one of the broad classes of coherence relations. Moreover, this coherence in cause and result is much stronger than the one in the general narrative due to the strong dialecticity and duality between the cause and result of causal narrative.

Inspired by these studies, we introduce the idea of narrative understanding into the ECE task. Specifically, considering that causal narrative involves two possibilities of chronological narrative and flashback narrative, our proposed approach improves the sequential semantics modeling in current studies and focuses on learning the two possible semantic information of causal narrative. Note that, this is the first work that uses causal narratives of discourse to address this task.

3 Problem Definition

The formal definition of the ECE task is listed below. The inputs are the texts of document D and an emotion phrase ep, where D represents a causal text with n clauses, and ep refers to the overall emotion of this document. Here, D is denoted as follows:

$$D = \{c_1, c_2, ..., c_n\}, \tag{1}$$

where c_i is the i^{th} clause in document D. And the emotion of ep is uniquely consistent with that of a clause in D. The goal of the ECE task is to learn a function F to identify the cause clauses in document D that trigger the emotion ep:

$$\boldsymbol{y} = F(D,\ ep), \tag{2}$$

here, $\boldsymbol{y} = \{y_1, y_2, ..., y_n\}$ denotes the emotion cause labels of the clauses in D, where y_i = 1 if the i^{th} clause is an emotion cause clause, else y_i = 0. In this paper, we regard the clause in D whose emotion is consistent with that of ep as the emotion result clause.

The above definition about ECE could be illustrated via the instance in Fig. 1. In terms of the form for the input texts, the inputs to be processed are a document containing multiple clauses and a phrase. As for the textual content, this instance is a causal text with an overall emotion category *"grateful"*. Based on the above, this article aims to reveal the clauses that trigger the emotion *"grateful"* of the document: if a clause triggers the overall emotion of its document, we will mark it as the emotion cause clause of this document; otherwise, we will mark it as a non-cause clause.

For ease of explanation, relevant notations used in this article are summarized in Table 1.

4 Method

In this section, we will introduce model details as well as model training of our proposed Causal Narrative Comprehension Model (CNCM).

4.1 Overall Architecture

The architecture of CNCM is shown in Fig. 2, which consists of: 1) *Input Embedding*: accomplishing the clause embeddings of the input texts; 2) *Emotion Causality Understanding*: acquiring emotional causal correlations in causal narratives via the REA unit; 3) *Cause Narrative Representation*: representing textual cause narratives through the NCA unit; 4) *Emotion Causality Re-understanding*: re-understanding emotional causal correlation under the guidance of cause narrative representation via the REA unit; 5) *Cause Prediction*: predicting the emotion cause label for each clause.

The process is summarized below. First, we mark the emotion result clause according to the cue of the given emotion phrase and realize the embedded representation of each clause for the document. Second, we adopt the REA

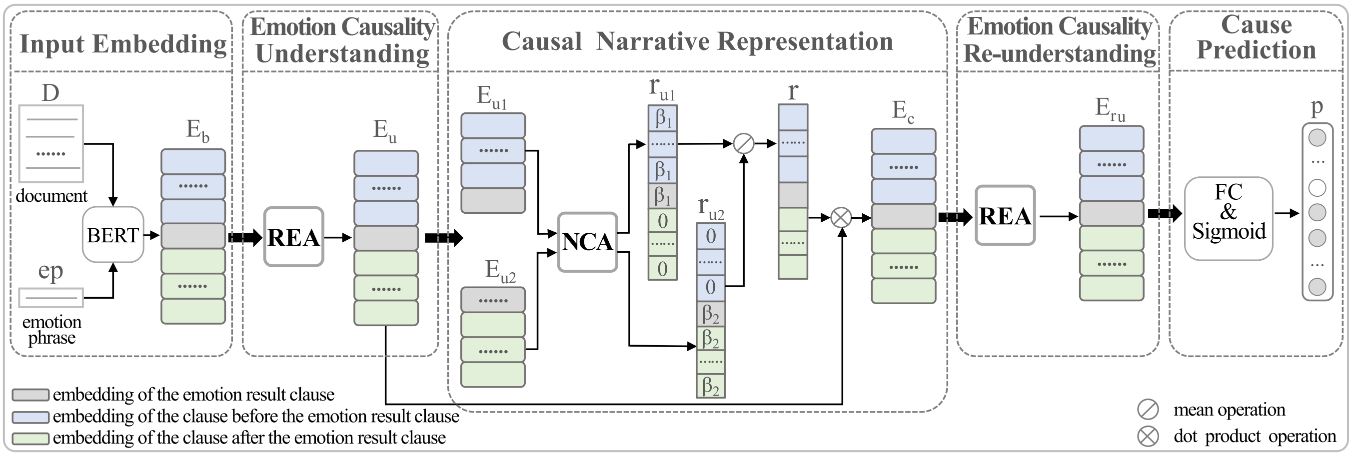


Fig. 2. The overall architecture of Causal Narrative Comprehension Model (CNCM) for emotion cause extraction.

unit to mine the emotional causal correlation between the known result clause and other clauses for the preliminary emotion causality understanding. Next, we perform the NCA unit, which utilizes the semantic coherence of causal narrative to learn the document's possible causal narrative for realizing the causal narrative representation of the document. Finally, imitating the human habit of repeated comprehension while reading long text, we execute the REA unit again to re-understand the emotional causality of the document. Based on the final effective text representations, we predict the emotion cause clause to realize this task. The details are introduced in the following parts.

4.2 Input Embedding

As stated in the problem definition above, to cognize the causal narrative of document D for the subsequent narrative comprehension, we need to mark the result clause at the first step. For our benchmark datasets, the English dataset has been annotated with the information about the emotion result clause, and the Chinese dataset can realize the location of the emotion result clause by finding phrase ep in document D through string matching operation. Afterward, the mathematical form of document D in Eq. (1) can be modified as below:

$$D = \{c_1, c_2, ..., c_{k-1}, c_k^e, c_{k+1}, ..., c_n\}, \tag{3}$$

where c_k^e is the emotion result clause containing the expression consistent with ep. After that, the input texts would be fed into the encoder for vector forms. Considering that the pre-trained language model BERT [77] has excellent capability in semantic representations, especially its evolution model BERT-wwm [78] achieves considerable performance in Chinese based on whole-word masking technology and a large-scale Chinese corpus. we adopt BERT-wwm to accomplish the vectorization of input texts:

$$\boldsymbol{E}_b = BERT(D) = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_{k-1}, \boldsymbol{x}_k^e, \boldsymbol{x}_{k+1}, ..., \boldsymbol{x}_n\}, \tag{4}$$

where $\boldsymbol{E}_b \in \mathbb{R}^{n \times d_w}$ represents the documental semantic embeddings at sentence level. d_w is the dimension of the output embeddings in BERT-wwm. $\boldsymbol{x}_i$ is the embedding of the i^{th} clause. Specially, x_i^e refers to the embedding of the result clause c_k^e when $i = k$.

4.3 Emotion Causality Understanding

Inspired by the human habit of first preferring an initial grasp when reading a complex text [79], at the beginning of CNCM, we also try to make a preliminary understanding of the text, especially its causal narrative. Currently, a plethora of studies have established the importance of correlations between causes and results in causal texts understanding [20], [21], [80], [81]. Hence, as depicted in Fig. 3, we focus on emotional causal correlations and utilize the known result clause c_k^e to design a Result-aware Emotional Attention (REA) unit to explore this emotional causal association for emotion causality understanding.

REA Unit. Considering the foundational role of clause representational quality in causal association modeling, we first focus on representing the clauses accurately before achieving the emotional causal association among these clauses. Since BiLSTM is good at modeling long texts and capturing context information, we employ it to cope with document D, ensuring that each clause' semantics would not deviate from the context information of D:

$$\boldsymbol{E}_h = BiLSTM(\boldsymbol{E}_b), \tag{5}$$

where $\boldsymbol{E}_h = \{\boldsymbol{h}_1, ..., \boldsymbol{h}_k^e, ..., \boldsymbol{h}_n\} \in \mathbb{R}^{n \times d_h}$ is the hidden state of each clause in document D processed by BiLSTM. d_h is the dimension of the hidden state in BiLSTM. $\boldsymbol{h}_k^e$ is the hidden state of the result clause. By this way, the context information of document D is incorporated into the representation of each clause.

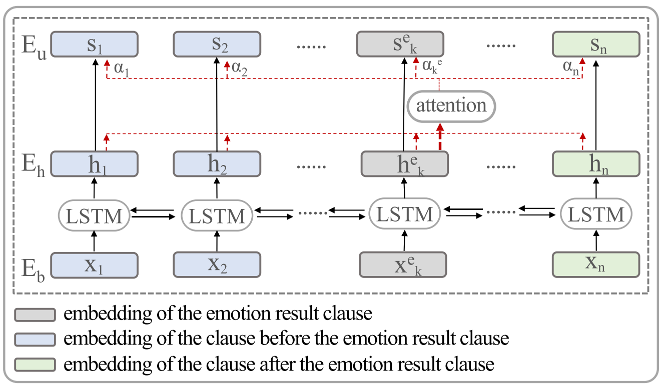


Fig. 3. The architecture of REA.

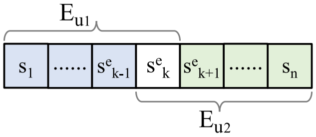


Fig. 4. The causal narrative in a document.

After acquiring the good representation by integrating document context, we construct the emotional causal correlation based on the known result clause as follows. So far, narratological researches have shown that sentences of a causal narrative varied in the extent to which the causal relations are close or distant [20], [80]. Specific to the clauses of causal narrative in document D, the ones that are more correlated with the result clause in semantics are more likely to be cause clauses. Considering that attention mechanism can simulate humans' visual attention and highlight critical information by concentrating on differences of its input, we adopt an attention mechanism to capture different causal associations between the result clause and each clause:

$$\begin{aligned} \boldsymbol{M} &= \tanh\big(\boldsymbol{W}_1\boldsymbol{E}_h + \boldsymbol{W}_2\boldsymbol{h}_k^e\big),\\ \boldsymbol{\alpha} &= softmax(\boldsymbol{W}_3 M), \end{aligned} \tag{6}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^n$ is the attention weight scores, standing for each clause's emotional causal association to the current representation $\boldsymbol{h}_k^e$ of the result clause. $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ and $\boldsymbol{W}_3$ are the trainable parameters. Afterward, $\boldsymbol{\alpha}$ is fused into the document representation $\boldsymbol{E}_h$, promoting CNCM focusing on the semantics more relevant to the emotion result:

$$\boldsymbol{E}_u = \boldsymbol{\alpha} \cdot \boldsymbol{E}_h, \tag{7}$$

where $\boldsymbol{E}_u = \{\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_k^e, ..., \boldsymbol{s}_n\} \in \mathbb{R}^{n\times d_h}$ is the text representation containing emotional causal association, which refers to the preliminary cognition of emotion causality of document D.

4.4 Causal Narrative Representation

To address the first challenge of our work stated before, this layer purposes to model the causal narrative information of document D. Currently, most of the narratological studies focus on modeling sequential semantics of clauses in documents [73]. However, these sequential semantics-based models are not suitable for causality modeling of causal texts since the cause and result clauses in causal texts are not always narrated sequentially. Fortunately, causal narratives contain special narrative properties, which could deal with this issue. According to narratology, causal narrative usually includes chronological narrative (cause–before–effect) and flashback narrative (effect–before–cause) [22], [28]. It implies the regions before or after the result in a causal text could be regarded as alternative cause regions of the text.

Special to document D, we leverage the known result clause $\boldsymbol{s}_k^e$ to speculate the alternative cause regions $\boldsymbol{E}_{u1}$ and $\boldsymbol{E}_{u2}$ of document D, as shown in Fig. 4. Namely, the regions $\boldsymbol{E}_{u1}$ and $\boldsymbol{E}_{u2}$, which are adjacent and contain the result clause $\boldsymbol{s}_k^e$, are two alternative cause regions of D:

$$\begin{aligned} \boldsymbol{E}_{u1} &= \{\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_k^e\},\\ \boldsymbol{E}_{u2} &= \{\boldsymbol{s}_k^e, \boldsymbol{s}_{k+1}, ..., \boldsymbol{s}_n\}. \end{aligned} \tag{8}$$

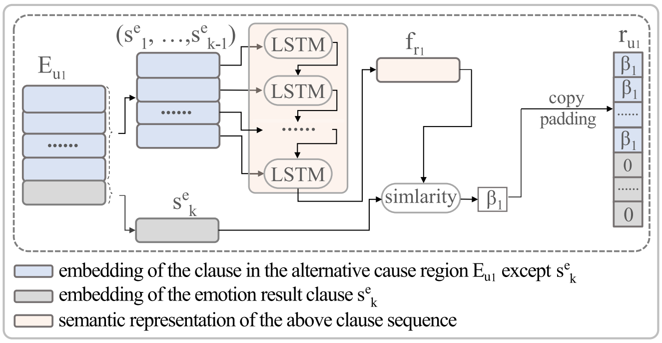


Fig. 5. The architecture of CNA. Here, we take the region $\boldsymbol{E}_{u1}$ as an example to illustrate its process.

This reveals that document D may contain two alternative causalities: one is the narrative between the region $\boldsymbol{E}_{u1}$ and the emotion result clause $\boldsymbol{s}_k^e$, and the other is the narrative between $\boldsymbol{E}_{u2}$ and $\boldsymbol{s}_k^e$.

According to the role of causality in narrative understanding [24], [25], the above awareness of causal narrative about document D can be helpful in determining its precise region of cause clauses. Thus, we firstly devise a narrative-aware causal association (NCA) unit to model the two possible causal narratives in regions $\boldsymbol{E}_{u1}$ and $\boldsymbol{E}_{u2}$ respectively. Subsequently, the causal narrative information of these two regions is integrated together for the causal narrative association vector r of its whole document D. To pursue effective text representations, we integrate the causal narrative information $\boldsymbol{r}$ into the document feature $\boldsymbol{E}_u$ from the previous layer:

$$\boldsymbol{E}_c = \boldsymbol{r} \cdot \boldsymbol{W}_4\boldsymbol{E}_u, \tag{9}$$

where $\boldsymbol{E}_c \in \mathbb{R}^{n\times d_m}$ refers the causal narrative representation of document D. d_m is the dimension of the hidden layer in CNCM. $\boldsymbol{W}_4$ is the trainable parameters. Guiding by causal narrative representation, CNCM can capture causal narrative information of ECE documents, which is conducive to emotion cause detection.

CNA Unit. Here, we take the alternative cause region $\boldsymbol{E}_{u1} : \{\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_k^e\}$ as an example to illustrate the implement process of CNA, as shown in Fig. 5. Considering that the clause $\boldsymbol{s}_k^e$ is the known result of region $\boldsymbol{E}_{u1}$, this region may involve a causal narrative. Namely, there may be a possible causal association between the sequence $[\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_{k-1}]$ and $\boldsymbol{s}_k^e$ of this region. Inspired by the semantic coherence in causal narratives [32], [82], this possible causal association is manifested by the semantic coherence between the two. Hence, we utilize the semantic correlation between the sequence $[\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_{k-1}]$ and result clause $\boldsymbol{s}_k^e$ to measure the possible causal association between the two for the causal narrative modeling of the region $\boldsymbol{E}_{u1}$.

First, according the semantic consistency within discourse [25], [30], [32], the semantics of the sequence $[\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_{k-1}]$ is can be represented by its subsequent semantics because the semantic consistent of the two. Specifically, considering that LSTM [16] is good at processing and understanding the semantics of sequence data, we utilize LSTM to handle the sequential clauses of this sequence by time step for the representation of this sequence:

$$\boldsymbol{f}_{r1} = LSTM([\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_{k-1}]), \tag{10}$$

where $\boldsymbol{f}_{r1} \in \mathbb{R}^{d_h}$ is the output of LSTM at its final step and can be regarded as the representation of $[\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_{k-1}]$. Second, we refer to the narrative modeling in the study about story completion [73] and adopt cosine similarity to figure out the semantic correlation between this sequence and the result clause $\boldsymbol{s}_k^e$:

$$\beta_1 = Similarity\big(\boldsymbol{f}_{r1}, \boldsymbol{s}_k^e\big) = \frac{\big(f_{r1} \times \boldsymbol{s}_k^e\big)}{\sqrt{\left(\boldsymbol{f}_{r1}\right)^2} \times \sqrt{\left(\boldsymbol{s}_k^e\right)^2}}, \tag{11}$$

where $\beta_1 \in \mathbb{R}$ is used to measure the degree of likelihood of causality in the alternative cause region $\boldsymbol{E}_{u1}$.

Similarly, we utilize the CNA unit to handle the other alternative cause region $\boldsymbol{E}_{u2}$ to acquire the value $\beta_2 \in \mathbb{R}$, which is used to measure the possibility of causal narrative in $\boldsymbol{E}_{u2}$. And for the sake of calculation, we convert β_1 and β_2 to the corresponding vector form by copying and padding operations, respectively:

$$\boldsymbol{r}_{ui} = \begin{cases} (\underbrace{\beta_1, \beta_1, ..., \beta_1}_{k}, \underbrace{0, 0, ..., 0}_{n-k}), & i = 1, \\ \\ (\underbrace{0, 0, ..., 0}_{k-1}, \underbrace{\beta_2, \beta_2, ..., \beta_2}_{n-k+1}), & i = 2, \end{cases} \tag{12}$$

where $\boldsymbol{r}_{ui} \in \mathbb{R}^n$ stands for the causal narrative association vector of the region $\boldsymbol{E}_{ui}$. k is the number of clauses in $\boldsymbol{E}_{u1}$, while $n-k+1$ is the number of clauses in $\boldsymbol{E}_{u2}$.

Moreover, considering that each discourse segment has both local and global coherence [30], $\boldsymbol{r}_{u1}$ and $\boldsymbol{r}_{u2}$ are just involve the possibilities of local causal narratives in document D. Thus, it is necessary to integrate the two together for the global causal narrative of this document. Since $\boldsymbol{r}_{u1}$ and $\boldsymbol{r}_{u2}$ contain the desired information at the position of clause $\boldsymbol{s}_k^e$, we realize this integration by the mean operation as below:

$$\boldsymbol{r} = (\boldsymbol{r}_{u1} + \boldsymbol{r}_{u2})/2, \tag{13}$$

where $\boldsymbol{r} \in \mathbb{R}^n$ refers to the causal narrative association vector of document D. As stated before, $\boldsymbol{r}$ is integrated into the document representation E_u to achieve the causal narrative modeling of this document and obtain the effective document representation E_c.

4.5 Emotion Causality Re-Understanding

Inspired by humans' multiple understanding of long texts [79], we in this section employ the REA unit again to re-understand the causality of document D. Specifically, the above text representation $\boldsymbol{E}_c$ is fed into REA as below:

$$\boldsymbol{E}_{ru} = REA(\boldsymbol{E}_c), \tag{14}$$

where $\boldsymbol{E}_{ru} \in \mathbb{R}^{n \times d_h}$ stands for the final text representation after this re-cognition of emotional causality of document D. It could facilitate CNCM to comprehend the emotional causality of this document very well. The relevant details are similar to those of the preliminary understanding of emotional causal correlation for document D.

4.6 Cause Prediction

In this layer, we leverage the final representation $\boldsymbol{E}_{ru}$ of document D to predict emotion cause clauses. Specifically, we fed each clause embedding of $\boldsymbol{E}_{ru}$ successively into a single-layer fully connected (FC) network and a sigmoid function for cause clauses prediction:

$$\boldsymbol{p} = Sigmoid(FC(\boldsymbol{E}_{ru})), \tag{15}$$

where $\boldsymbol{p} \in \mathbb{R}^n$ indicates the probability vector of emotion cause labels for the clauses in document D.

4.7 Model Learning

Since the ECE task is a classification problem, we employ the *cross–entropy* function as our loss function.

$$Loss = -\frac{1}{m}\frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n} y_i^j \log p_i^j, \tag{16}$$

where m indicates the number of documents in the datasets. n means the number of clauses in a random document. y_i^j refers to the true cause label of the clause c_i in the j^{th} document of the datasets. To minimize the *loss*, we use the Adam optimizer to update the parameters of each layer. Additionally, we conduct the dropout operation and K-fold (k=10) cross validation trick to prevent our model training from overfitting.

5 Experiments Study

In this section, we first introduce the experiment preparation, involving the dataset details, evaluative criteria, and experimental settings. Then, we list some baseline methods, analyze the experimental results in detail, and conduct some ablation studies. Subsequently, we make some detailed analysis of some meaningful issues for our proposed model. Finally, we present several visualization cases to illustrate the workflow of CNCM.

5.1 Dataset Description

Following the general practice in many previous studies [4], [14], we conduct experiments on two publicly ECE corpuses: a Chinese benchmark dataset [48] based on Sina City News [2] and an English benchmark dataset [3], [19], [83] based on an English novel. To provide an intuitive sense of the datasets, we also present a document example from the Chinese dataset as shown in Fig. 6. It contains 5 clauses and is annotated with an emotion phrase. Additionally, we present some key information about these two datasets in Table 2. Particularly, most of the documents in both datasets contain one cause clause, accounting for more than 90.20% of the total. In contrast, merely a few documents contain 2 or more cause clauses. Noticeably, according to the numbers of the cause clause and general clause (the non-cause clause) shown in Table 2, we can conclude that the size of each dataset is not large enough and the number of cause clauses in documents is extremely uneven.

Given that document size is closely related to document narrative complexity, we perform statistical analysis on the

2. https://city.sina.com.cn/

• 确认没有减少后，	• After confirming no reduction,
• 将钱包交还给温某。	• he returned the wallet to Wen.
• 钱包短短半小时失而复得，	• Wallet was lost and recovered in just half an hour,
• 失主感到十分庆幸，	• the owner felt very **fortunate**,
• 连连对民警表示感谢。	• and repeatedly thanked the police.
• **庆幸**	• `fortunate`

Fig. 6. A document example in the Chinese dataset. The original text is on the left, while the corresponding English translation for each clause is listed on the right.

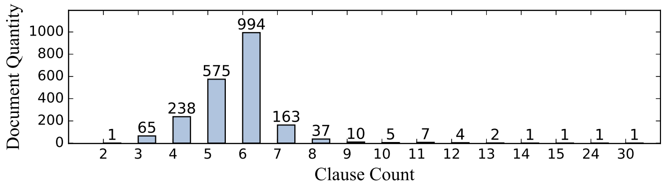


Fig. 7. The statistics of document size by clause count in the Chinese dataset.

document size of the two datasets according to the number of clauses, shown in Figs. 7 and 8. It is can be observed that the document size distribution of the Chinese dataset is relatively centralized, while that of the English dataset is very scattered. Specifically, the clause count of documents in the Chinese dataset is mainly distributed from 3 to 11. In contrast, the clause count of documents in the English dataset is mainly distributed from 1 to 25. In short, the document size of the Chinese dataset is relatively simple and moderate compared to that of the English dataset.

5.2 Evaluative Criteria

Following [14], [19], [52], we also adopt Precision, Recall, and F1 to evaluate the performance of models in this paper.

$$Precision = \frac{\sum_{c\_cause} 1}{\sum_{p\_cause} 1}, Recall = \frac{\sum_{c\_cause} 1}{\sum_{a\_cause} 1}, F1 = \frac{2 * Precision * Recall}{Precision + Recall},$$

where $\sum_{c\_cause}$ is the predicted correct cause clause. $\sum_{p\_cause}$ means all the predicted cause clauses. $\sum_{a\_cause}$ stands for the annotated cause clauses.

5.3 Experimental Settings

For better training, we initialize the related parameters of our model by the following settings. First, we split the dataset into training (80%) and test (20%) sets. In terms of data vectorization, consistent with [18], we adopt the advanced pre-trained language model BERT [78] to encode each input clause into an embedding vector. By investigation of BERT-wwm used in other works, we fine-tune the pre-trained model on the training set and acquire the text vectors with the dimension d_w = 1,024. During training, the initial weights are assigned according to the uniform distribution suggested in [84]. During intermediate steps, we set d_h = 512 to be the dimensions of hidden state in the LSTM and BiLSTM of CNCM. Empirically, the dimension d_m of ordinary hidden layers in CNCM is set to 64 to ensure the consistency of the dimension of text representation. Besides, we use the Adam optimizer with the learning rate of 0.005 to train the networks on an NVIDIA Tesla K80 GPU with the batch size of 128. And an early stop strategy is employed to stop the training process when the validated indicators fail to improve after 10,000 loops. Considering that the datasets are not very large and the number of cause clauses in documents is extremely uneven, we adopt 10-folds cross-validation and the dropout rate of 0.5 to mitigate possible overfitting.

TABLE 2
Key Information about the Datasets

Item	Number	Percentage
Chinese Dataset		
Documents	2,105	-
Clauses	11,799	-
Emotion cause clauses	2,167	-
Documents with 1 cause clause	2,046	97.20%
Documents with 2 cause clauses	56	2.66%
Documents with 3 cause clauses and more	3	0.14%
English Dataset		
Documents	2,156	-
Clauses	13, 838	-
Emotion cause clauses	2,421	-
Documents with 1 cause clause	1,949	90.40%
Documents with 2 cause clauses	164	7.61%
Documents with 3 cause clauses and more	43	2.00%

5.4 Baseline Methods

We compare our proposed CNCM with the following groups of baselines:

- *Rule-based methods: CB* [46] is also a traditional method that introduces commonsense knowledge into the ECE task to reveal emotion causes. *RB* [11] is an original ECE method that is based on manual rules for emotion cause detection.
- *Machine learning-based methods: SVM+word2vec*[85] combines the strength of SVM classifier and word2-vec embedding [86] to recognize cause clauses. *SVM +RB+CB* [83] is an SVM framework [87] fused with rules and knowledge for emotion cause identification. *Multi-kernel* [48] is a machine learning approach which employs the multi-kernel classifier for emotion cause extraction. *LambdaMART* [53] transforms the ECE task to a ranking problem and selects cause clauses by rank criterion.

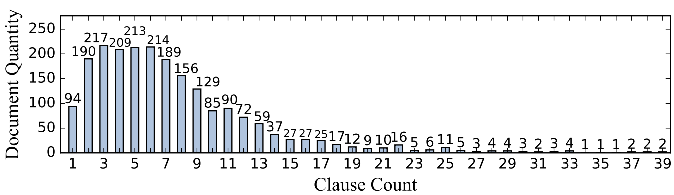


Fig. 8. The statistics of document size by clause count in the English dataset.

TABLE 3
Performances of Different Models on the Chinese Dataset

Model	Precision	Recall	F1
(1) CB [46]	26.72%	71.30%	38.87%
(2) RB [11]	67.47%	42.87%	52.43%
(3) SVM+word2vec [85]	43.01%	42.33%	41.36%
(4) SVM+RB+CB [83]	59.21%	53.07%	55.97%
(5) Multi-kernel [48]	65.88%	69.27%	67.52%
(6) LambdaMART [53]	77.20%	74.99%	76.08%
(7) Memnet [52]	70.76%	68.38%	69.55%
(8) CANN [50]	77.21%	68.91%	72.66%
(9) HCS [13]	73.88%	71.54%	72.69%
(10) MANN [15]	78.43%	75.87%	77.06%
(11) FSS-GCN [4]	78.61%	75.72%	77.14%
(12) EF-BHA [54]	79.38%	78.08%	78.68%
(13) RHNN [14]	81.12%	77.25%	79.14%
(14) CNCM	**93.97%**	**78.85%**	**85.75%**

TABLE 4
Performances of Different Models on the English Dataset

Model	Precision	Recall	F1
(1) Memnet [52]	46.05%	41.77%	43.81%
(2) MANN [15]	**79.33%**	40.81%	53.28%
(3) FSS-GCN [4]	67.43%	53.03%	59.48%
(4) RHNN [14]	69.01%	52.67%	59.75%
(5) EF-BHA [54]	72.77%	53.05%	**61.37%**
(6) CNCM	57.69%	**62.10%**	59.79%

- *Deep learning-based methods: Memnet* [52] designs a question answering framework which understands emotion information well by memory network to address the ECE task. *CANN* [50] employs attention mechanism to enhance text representations by capturing the mutual impacts between the emotion clause and the other clauses for the ECE task. *HCS* [13] develops a hierarchical network-based clause selection framework to model multi-granularity semantic features to identify emotion cause. *MANN* [15] is a multi-attention-based network which realizes the emotion cause attention and candidate clause attention for good text understanding and emotion cause extraction. *FSS-GCN* [4] is a graph convolutional networks incorporating text semantics and structural information. *EF-BHA* [18] models the document-level context and mines clause relations based on emotion clauses the ECE task. *RHNN* [14] is a novel hierarchical neural model that fuses the sentiment lexicon and common knowledge via restraining parameters for the ECE task.

5.5 Effectiveness Comparison

The experimental results on the Chinese and English datasets are demonstrated in Tables 3 and 4. Here, we analyze these two tables, respectively.

Performance on the Chinese Dataset. Table 3 shows the effectiveness comparison related to the Chinese dataset. It can be found that the Rule-based methods perform relatively poorly. It is probably because they rely heavily on linguistic rules or commonsense, but manual rules or knowledge can not cover all complex linguistic phenomenons. Besides, the indicators of the CB method differ substantially. It may imply this approach requires more appropriate commonsense knowledge. In contrast, machine learning-based approaches can better mine the information features of text corpuses, so the corresponding performances are improved to some extent. However, the performance gains of such methods are limited because they rely on complex artificial features. Evidently, the performances of deep learning-based methods have generally been further improved. It is probably because deep neural networks can simulate human brains to understand semantics very well. Among these methods, the results of RHNN and EF-BHA are quite competitive. RHNN incorporates discourse context and prior knowledge, which can contribute to emotion semantics understanding. While EF-BHA adopts the boundaries detecting method of emotion cause spans, which can reduce the range of detection clauses and improve the efficiency of the model. Noticeably, CNCM achieves a performance improvement of 6.6% about the F1 score than the advanced RHNN model. The superiority of CNCM may attribute to its grasp of causal narrative in documents, which enables CNCM to understand the overall emotional causality of the document more directly and accurately.

Performance on the English Dataset. The performance comparison related to the English dataset is shown in Table 4. It can be observed that the indicators of these approaches in Table 4 are generally inferior to their indicators in Table 3. And so does our model CNCM. That is, the models do not perform as well on the English dataset as they do on the Chinese dataset. It may be due to textual differences between the two datasets. As mentioned before, the Chinese and English datasets are derived from a news corpus and an English novel respectively. Compared with the Chinese dataset, the writing style of text in the English dataset is relatively free and literary, which leads to its relatively complicated semantics and narratives. This may account for the generally poor performance of models in English datasets. Moreover, as demonstrated in Section 5.1, the distribution of document sizes in the English dataset is much more complex than that in the Chinese dataset. It indicates a greater diversity of document structures in the English dataset and also implies more complex narratives in this dataset. Obviously, this would hinder the performance of CNCM. Because CNCM is based on a causal narrative understanding and relatively sensitive to the narrative characteristics of texts. In summary, the characteristics of language expression and document structure in the English dataset together lead to performance degradation of CNCM. As shown in Table 4, although the overall performance of CNCM is slightly weaker than that of the most advanced model EF-BHA, it is comparable to that of the suboptimal model RHNN.

Comprehensive Comparison. Combining Tables 3 and 4 together, we can find that EF-BHA and RHNN are the SOTA benchmark models. Furthermore, it can be observed that these two SOTA benchmark models exhibit different performance characteristics compared with CNCM on the Chinese dataset and the English dataset, which are in line with the analysis above. Thus, to scientifically evaluate whether our CNCM model is superior to existing models,

TABLE 5
Ablation Performances of CNCM

Model	Precision	Recall	F1
(1) CNCM (w/o NCA)	90.22%	74.25%	80.89%
(2) CNCM (w/o BERT)	89.71%	80.38%	84.75%
(3) CNCM (w/o REA_1)	82.72%	**82.53%**	82.62%
(4) CNCM (w/o REA_2)	87.11%	77.70%	82.14%
(5) CNCM	**93.97%**	78.85%	**85.75%**

we perform holistic t-tests of the overall performance for CNCM and these two SOTA benchmark models (EF-BHA and RHNN). Through the t-tests, we can acquire over 95%, and 99% of confidence that CNCM has significant improvement over EF-BHA and RHNN, respectively, which indicates that CNCM has certain superiority.

5.6 Ablation Study

To confirm the effectiveness of each unit in CNCM, we conduct ablation experiments via removing or replacing the following three aspects: the causal narrative representations of NCA, the pre-training model for text embeddings, and the context-aware emotion attention of REA. The related results are shown in Table 5.

5.6.1 *Causal Narrative Representation of NCA*

As stated above, the essential innovation of our proposed model is the causal narrative representations of the NCA unit. To examine the effect of this innovation on the final experimental performance, we conduct experiments on the ablation model CNCM (w/o NCA), which is derived from CNCM by removing the NCA unit. As shown in Table 5, CNCM (w/o NCA) without causal narrative representations has the worst performance. It suggests that the causal narrative representations of causal texts could greatly contribute to the performance improvement of the ECE task. It might be because the causal narrative representation of the NCA unit could guide CNCM to focus on the emotion cause region which is more related to the emotion result clause.

5.6.2 *Pre-Trained Model for Text Embedding*

As we know, BERT can output outstanding text representations which have been shown to be very effective in many downstream tasks of natural language processing. Owing to this reason, we achieve the initial vectorization of texts in CNCM through BERT's evolution model BERT-wwm. To explore the importance of the pre-trained language model on the ECE task, we replace BERT-wwm with another popular pre-trained language model, word2vec [86] to conduct experiments. The corresponding ablation model is CNCM (w/o BERT). As can be seen from Table 5, the performance degradation of CNCM (w/o BERT) is very small. It seems to indicate the superiority of CNCM may not be mainly attributed to BERT-wwm.

Besides, it can be observed that CNCM (w/o BERT) performs quite well without BERT. For one thing, this may benefit from the contribution of the NCA unit. As shown in Table 5, CNCM (w/o BERT) has the smallest performance degradation, while CNCM (w/o NCA) has the largest performance degradation. It demonstrates that the NCA unit might contribute more to the effectiveness of CNCM than BERT. For another, due to the nature of high complexity and lack of large scale corpora, it is usually difficult to train BERT-based models on small scale datasets of the ECE task. Therefore, our proposed CNCM can still perform well on the ECE task without BERT.

5.6.3 *Causal Association Cognition of REA*

Inspired by the reading habits of human beings [79], CNCM twice utilizes the REA unit to understand the emotional causal association of causal narrative. Actually, the REA unit in the layer of emotion causality understanding is designed to obtain preliminary cognition about the emotional causal association of documents. By comparison, the REA in the layer of emotion causality re-understanding aims to better understand the emotional causal association with the guidance of the preliminary cognition and the causal narrative representations of documents. The proposed ablation models are the model CNCM (w/o REA_1) and CNCM (w/o REA_2), where CNCM (w/o REA_1) is the evolution model without the REA unit of the preliminary understanding phrase, and CNCM (w/o REA_2) is the one without the REA unit of the emotion causality re-understanding phrase. As shown in Table 5, after removing one REA unit respectively, these two evolution models achieve lower performance than CNCM. It implies that the REA unit has a great influence on our model.

Another meaningful observation in Table 5 is that the performance degradation of CNCM (w/o REA_2) is higher than the one of CNCM (w/o REA_1). It demonstrates that the second REA unit is something more important than the first one. Perhaps because the second REA unit is in the last stage of CNCM, while the first REA is in the initial stage of CNCM. Thus, the second REA unit has a more direct effect on CNCM than the first REA.

5.7 Detailed Analysis

In this subsection, we carry out some supplementary experiments to give a detailed analysis of our proposed model from multiple aspects.

5.7.1 *Effects of Cause-Result Order*

According to studies about narratology [22], [28], the cause-result order of causal narrative is either chronological narrative (cause–before–effect) or flashback narrative (effect–before–cause). Especially in our Chinese dataset, as shown in Table 6, the former and the latter account for 65.75% and 34.25%. This imbalance may be due to the fact that events typically are presented in chronological order in narratives [88]. Because this is consistent with the laws of events development.

Furthermore, we discuss the effects of the balance about cause-result order on the performance of the ECE task. As the results demonstrated in Fig. 9, "the ratio" denotes the ratio of chronological narrative and flashback narrative in the corresponding experiment. While "Rate_Actual" refers to this ratio in the original Chinese dataset. Obviously, CNCM performs best when the contents of ECE documents all conform to the chronological narrative. It may indicate

TABLE 6
Statistics of Cause-Result Order in the Chinese Dataset

Item	Number	Percentage
Documents with chronological narrative	1,384	65.75%
Documents with flashback narrative	721	34.25%
All documents	2105	100.00%

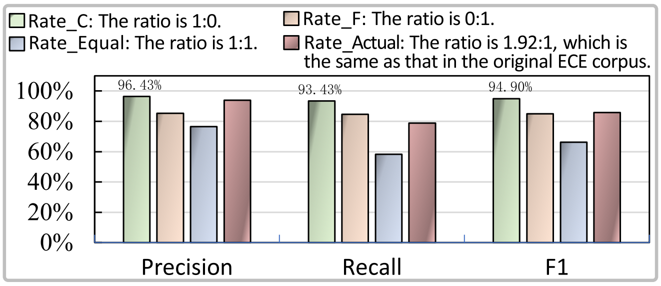


Fig. 9. The effects of cause-result order on the performance of CNCM. Here, the term "ratio" denotes the ratio of chronological narrative (cause–before–effect) and flashback narrative (effect–before–cause) in the Chinese dataset.

that CNCM is better at understanding texts with chronological narrative. This perhaps has something with the fact that chronological order is more consistent with the general norms and basic organization principles of narratives [88].

Significantly, balanced data usually has a positive effect on experimental performance. Yet, as shown in Fig. 9, CNCM performs worst when there is an equal proportion of chronological narrative and flashback narrative. This anomaly may be because the complex narrative in texts would increase the difficulty of semantic understanding. Comparatively speaking, the other experiments perform better than the one with balanced data. The reason may be that text semantics is easier to learn and master when the narrative mode of the corpus is relatively simple. The analysis of this anomaly may also be useful for the semantic understanding of other long texts. Namely, the simpler the linguistic patterns of the text are, the easier it is to learn.

5.7.2 Effects of Clause Distance

It is worthwhile to mention that the distance between emotion result clauses and cause clauses is one of the significant attributes of causal structure. Thus, we conduct relevant data statistics and experiments to explore the effects of this factor on the performance of CNCM.

Fig. 10 shows the distance statistics of emotion cause clauses relative to their emotion result clauses in the Chinese dataset. It can be observed that the positions of emotion cause clauses relative to the emotion result clauses within a document are usually no more than 3 clauses. Based on this observation, we choose the documents whose emotion cause clauses are no more than 1 or 2 or 3 clauses away from the emotion result clauses to construct three sub-datasets. The performance of CNCM on these 3 datasets is shown in Table 7. In this table, the term "Radius_1" indicates the document set, in which the distance of emotion cause clauses relative to emotion result clauses is no more than 1 clause. The same principle goes for the term

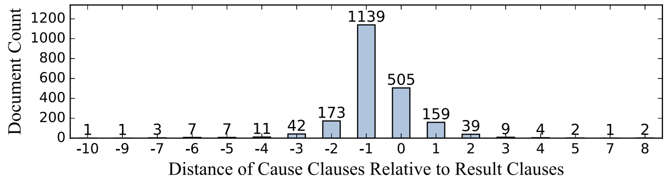


Fig. 10. The distance statistics of cause clauses relative to result clauses in the Chinese dataset. Here, the positive distance indicates that the cause clause comes before the result clause, while the negative distance is the opposite. If a document contains multiple cause clauses, we select its first cause clause for statistics.

TABLE 7
Performances of CNCM on Datasets with Different Distances of Emotion Cause Clauses Relative to Emotion Result Clauses

Dataset	Precision	Recall	F1
(1) Radius_1	93.00%	**89.12%**	**90.97%**
(2) Radius_2	92.25%	87.56%	89.75%
(3) Radius_3	91.89%	80.86%	85.91%
(4) Original Dataset	**93.97%**	78.85%	85.75%

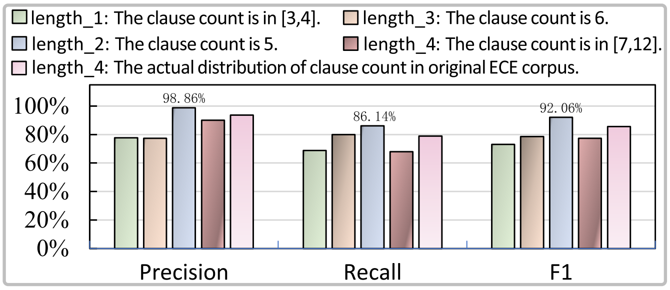


Fig. 11. The effects of document size on the performance of CNCM. Here, the document size refers to the clause count of documents in the Chinese dataset.

"Radius_2" and "Radius_3". Obviously, the data involved in "Radius_3" covers the vast majority of documents in the Chinese dataset, which is not difficult to explain why the performance of the "Radius_3" shown in Table 7 is close to that of the "Original Dataset". Additionally, the table also shows that the closer the distance between emotion cause clauses and emotion result clauses, the better the performance of CNCM. It may be due to that the closer distance between cause and result in a document, the tighter the causal semantics of the document. While the tight causal semantics implies that there is less information loss in the process of modeling causal structures. This would help to understand causal associations and identify emotion causes.

5.7.3 Effects of Document Size

Considering that document size is one of the important factors in determining the narrative complexity of a document, we also explore the effects of document size on CNCM. According the statistics of document size by clause count in Fig. 7, we select the document size in four ranges, i.e,3-4, 5, 6, 7-12 to conduct experiments. The corresponding results are shown in Fig. 11. Notably, CNCM performs best on all evaluation indicators when the document size amounts to 6 clauses. When the document size is less than or greater than 6, the results on all evaluation indicators correspondingly

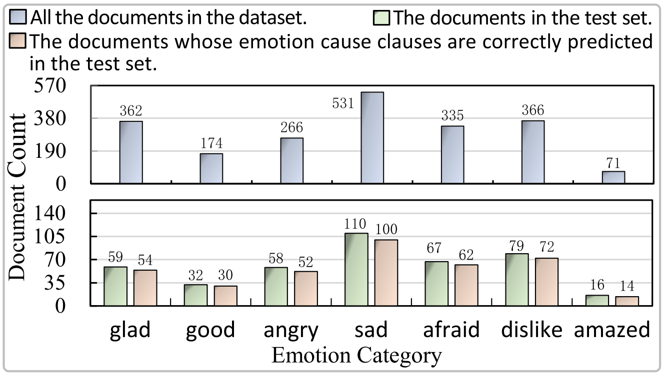


Fig. 12. The statistics of emotion category in the Chinese dataset.

decrease. It seems that moderate document size could help to document understanding. On one hand, the few clauses the document contains, the semantic information is less. Evidently, this case is not conducive to emotion cause identification of the ECE task because there is less information available for document understanding. On the other hand, the more clauses a document contains, the more complex its semantics and narrative become. In this case, document understanding becomes difficult, which may lead to the performance degradation of the ECE task.

Also note that the set of best results in Fig. 11 is even better than the results under the actual distribution of document size in the original dataset. It may be because various document sizes lead to complex narratives, which increase the difficulty of emotional semantics understanding. Furthermore, although the training data size of the former is smaller than that of the latter, the performance of the former is better than that of the latter. This is consistent with the findings in the previous analysis that the amount of training data does not significantly affect the performance of CNCM.

5.7.4 Effects of Emotion Category

In order to study whether CNCM had a bias for emotion categories when performing the ECE task, we also undertake a statistical analysis of emotion category. Taking the Chinese dataset as an example, first, we use the dictionary-based approach to identify the emotion of the documents in the dataset. Here, we use the Chinese emotion ontology database [89] to classify the emotions of these documents into seven categories: “glad”, “good”, “angry”, “sad”, “afraid”, “bad” and “amazed”. Second, we conduct statistics of emotion category for documents in the Chinese dataset and its test set, respectively. Finally, we perform the same statistics for the documents whose emotion cause clauses are correctly predicted in the test set. The statistical results are shown in Fig. 12. It can be remarked that there is little difference in the distribution of emotion categories of documents on the three datasets. This may suggest that our proposed model is insensitive to emotion categories.

5.7.5 Effects of Training Dataset Scale

To present the performance of CNCM systematically, we compare the results under different scales of the training dataset for our developed model. Taking the Chinese dataset as an example, the corresponding performances are shown in Table 8. Considering that the training dataset of the original experiment accounts for 80% of the total data in the Chinese dataset, we take into account the other three training data settings: 20%, 40%, and 60% of the total data. To be specific, we implement experiments under these training data settings and compare the corresponding performances to the original experiment. It can be found from Table 8 that the experimental performances of CNCM under different scales of the training dataset have little difference. These findings could indicate that our proposed model is still effective in the case of small training data.

TABLE 8
Performances of CNCM under Different Scales of Training Dataset of the Chinese Dataset

Training data size	Precision	Recall	F1
20% of Original Dataset	**92.63%**	76.12%	83.57%
40% of Original Dataset	92.48%	77.62%	84.40%
60% of Original Dataset	93.11%	**79.29%**	85.64%
80% of Original Dataset	93.97%	78.85%	**85.75%**

5.8 Case Studies

To provide some intuitive demonstrations of how causal narrative representation and emotional causal association improve the effectiveness of our model, we show some case studies in Fig. 13 to interpret what is happening in the working flow of CNCM. Fig. 13 A presents a correct instance, whose predicted cause label is the same as the truth label. In particular, as shown by the causal narrative representation of CNCM in the third column of this figure, the weight of the first alternative cause region is greater than that of the second region. It suggests the cause clause may be located in the first alternative cause region. This inference about the cause region is consistent with the ground truth. Moreover, in the fourth column of Fig. 13 A, the data highlighted by color represent the distribution of emotional attention of the last REA unit. Evidently, the distribution of emotional attention at clause 4 is larger than others. It can be conjectured that clause 4 should be the cause clause since it is most relevant to the result clause than other clauses. The above inferences are consistent with the ground truth, which illustrates CNCM can effectively locate cause regions and focus on clauses that are more relevant to emotion result clauses. Similarly, there is another instance shown in Fig. 13 B to illustrate the working details of CNCM.

In addition, we also provide two error cases to illustrate the existing problems of CNCM. As shown in the third column of Fig. 13 C, although our model could identify the accurate causal region by the learned causal narrative representation, the emotional causal association learned from the final emotional attention is biased. Therefore, CNCM misidentifies the adjacent non-cause clause of the real cause as a cause clause. Notably, the incorrect prediction sentence is located very close to the actual cause clause. It indicates that CNCM tends to be misled by the adjacent clauses of the true cause clause. As a further exploration, we aim to address this issue in the future.

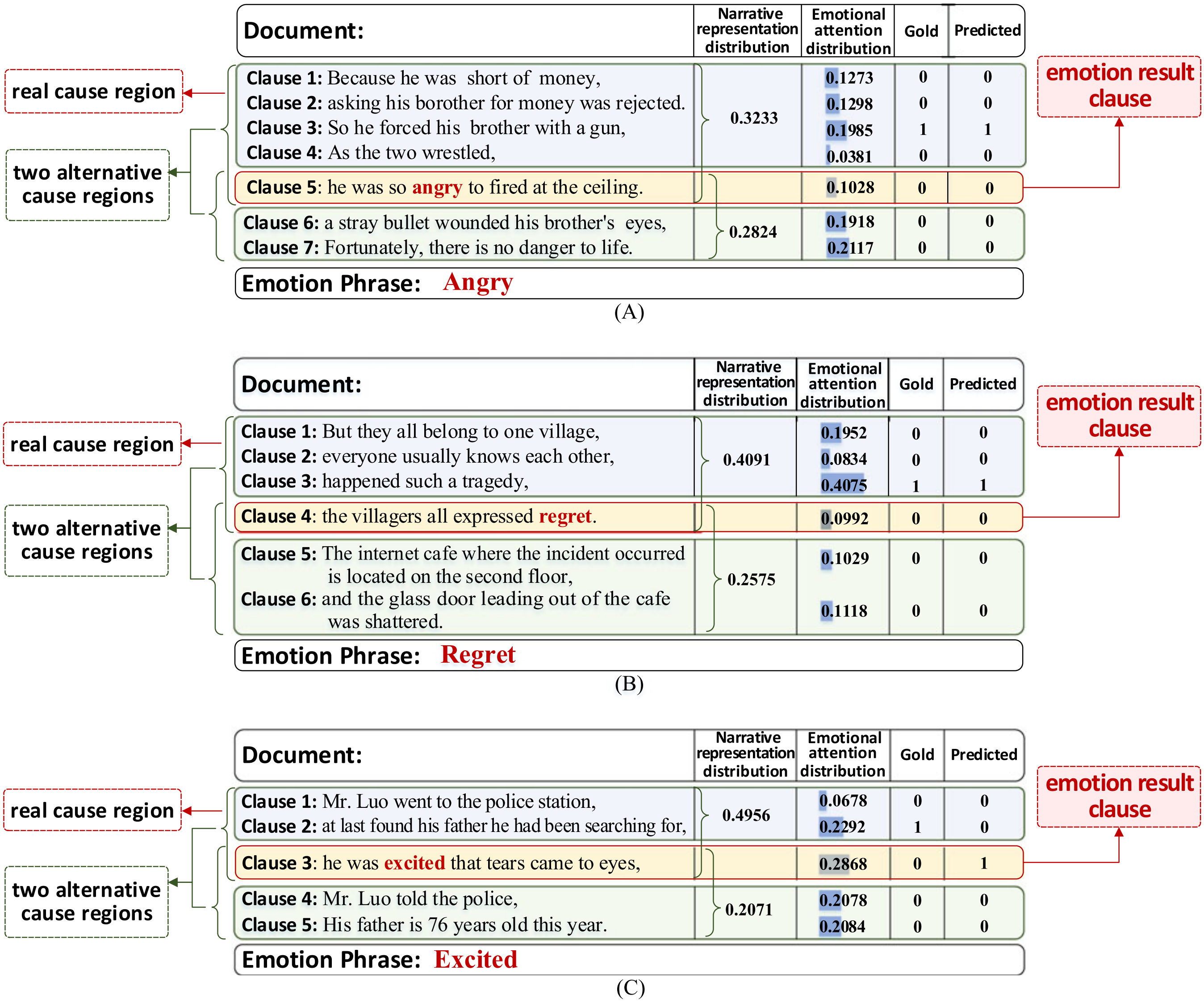


Fig. 13. Some examples of emotion cause extraction by CNCM. Figure (A) and (B) show correct cases, while figure (C) illustrates an error case.

6 Conclusion and Future Work

In this work, we focused on emotion cause extraction and argued that causal narrative comprehension is very important to this issue. To this end, we proposed a novel Causal Narrative Comprehension Model (CNCM) based on Causal Narrative Comprehension to address this task. Different from the previous works focusing on semantic understanding of clauses and emotion phrases, our proposed model focused on modeling and utilizing the causal narratives of documents to learn emotional causal association among clauses for emotion cause extraction. Specifically, CNCM utilized causal narrative to define cause regions and obtained causal narrative representations based on narrative coherence of the causes and the results. Guided by the representation of causal narrative, we developed a result-aware emotional attention unit to understand the emotional causal association multiple times, so as to realize the task of emotion cause detection. Extensive experimental results on the benchmark datasets demonstrated the effectiveness of CNCM for the ECE task.

In the future, we will strive to develop a multilingual corpus of the ECE task to refine our studies. Based on this, we also hope to conduct further research on much more general narrative material and attempt to make utilization of the narrative information to promote some appropriate tasks about text semantic understanding.

References

[1] S. Ruan et al., “Context-aware generation-based net for multi-label visual emotion recognition,” in *Proc. IEEE Int. Conf. Multimedia Expo*, 2020, pp. 1–6.

[2] Z. Ding, H. He, M. Zhang, and R. Xia, “From independent prediction to reordered prediction: Integrating relative position and global label information to emotion cause identification,” in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 6343–6350.

[3] Q. Gao et al., “Overview of ntcir-13 ECA task,” in *Proc. 13th NII Testbeds Community Informat. Access Res.*, 2017, pp. 165–192.

[4] G. Hu, G. Lu, and Y. Zhao, “FSS-GCN: A graph convolutional networks with fusion of semantic and structure for emotion cause analysis,” *Knowl.-Based Syst.*, vol. 212, 2021, Art. no. 106584.

[5] S. Ruan, K. Zhang, L. Wu, T. Xu, Q. Liu, and E. Chen, “Color enhanced cross correlation net for image sentiment analysis,” *IEEE Trans. Multimedia*, to be published, doi: 10.1109/TMM.2021.3118208.

[6] K. Zhang, L. Wu, G. Lv, M. Wang, E. Chen, and S. Ruan, "Making the relation matters: Relation of relation learning network for sentence semantic matching," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 14 411–14 419.

[7] J. Hu, Y. Liu, J. Zhao, and Q. Jin, "MMGCN: Multimodal fusion via deep graph convolution network for emotion recognition in conversation," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process.*, 2021, pp. 5666–5675.

[8] D. Li et al., "Enhancing emotion inference in conversations with commonsense knowledge," *Knowl.-Based Syst.*, vol. 232, 2021, Art. no. 107449.

[9] Y. Chen, S. Y. M. Lee, S. Li, and C.-R. Huang, "Emotion cause detection with linguistic constructions," in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 179–187.

[10] L. Gui, L. Yuan, R. Xu, B. Liu, Q. Lu, and Y. Zhou, "Emotion cause detection with linguistic construction in chinese weibo text," in *Proc. CCF Int. Conf. Natural Lang. Process. Chin. Comput.*, 2014, pp. 457–464.

[11] S. Y. M. Lee, Y. Chen, and C.-R. Huang, "A text-driven rule-based system for emotion cause detection," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 45–53.

[12] D. Ghazi, D. Inkpen, and S. Szpakowicz, "Detecting emotion stimuli in emotion-bearing sentences," in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*, 2015, pp. 152–165.

[13] X. Yu, W. Rong, Z. Zhang, Y. Ouyang, and Z. Xiong, "Multiple level hierarchical network-based clause selection for emotion cause extraction," *IEEE Access*, vol. 7, pp. 9071–9079, 2019.

[14] R. Xia, M. Zhang, and Z. Ding, "RTHN: A RNN-transformer hierarchical network for emotion cause extraction," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 5285–5291.

[15] X. Li, S. Feng, D. Wang, and Y. Zhang, "Context-aware emotion cause analysis with multi-attention-based neural network," *Knowl.-Based Syst.*, vol. 174, pp. 205–218, 2019.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[17] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

[18] X. Li, W. Gao, S. Feng, Y. Zhang, and D. Wang, "Boundary detection with bert for span-level emotion cause analysis," in *Proc. Findings Assoc. Comput. Linguistics*, 2021, pp. 1–6.

[19] C. Fan et al., "A knowledge regularized hierarchical approach for emotion cause analysis," in *Proc. Conf. Empir. Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process.*, 2019, pp. 5614–5624.

[20] J. M. Keenan, S. D. Baillet, and P. Brown, "The effects of causal cohesion on comprehension and memory," *J. Verbal Learn. Verbal Behav.*, vol. 23, no. 2, pp. 115–126, 1984.

[21] S. E. Pickren, M. Stacy, S. N. Del Tufo, M. Spencer, and L. E. Cutting, "The contribution of text characteristics to reading comprehension: Investigating the influence of text emotionality," *Reading Res. Quart.*, vol. 57, pp. 649–667, 2021.

[22] E. Kaiser, "Order of mention in causal sequences: Talking about cause and effect in narratives and warning signs," *Discourse Processes*, vol. 56, no. 8, pp. 599–618, 2019.

[23] T. Sanders, "Coherence, causality and cognitive complexity in discourse," in *Proc. 1st Int. Symp. Exploration Modelling Meaning*, 2005, pp. 105–114.

[24] T. Sanders and W. Spooren, "The cognition of discourse coherence," *Discourse Course*. Amsterdam, The Netherlands: John Benjamins, 2009, pp. 197–212.

[25] M. B. Wolfe, J. P. Magliano, and B. Larsen, "Causal and semantic relatedness in discourse understanding and representation," *Discourse Processes*, vol. 39, no. 2/3, pp. 165–187, 2005.

[26] P. Van den Broek, "The causal inference maker: Towards a process model of inference generation in text comprehension," *Comprehension Processes in Reading*. Mahwah, NJ, USA: L. Erlbaum, 1990, pp. 423–445.

[27] C. R. Fletcher and C. P. Bloom, "Causal reasoning in the comprehension of simple narrative texts," *J. Memory Lang.*, vol. 27, no. 3, pp. 235–244, 1988.

[28] T. Hoffmann, "Construction grammar as cognitive structuralism: The interaction of constructional networks and processing in the diachronic evolution of english comparative correlatives," *English Lang. Linguistics*, vol. 21, no. 2, pp. 349–373, 2017.

[29] T. Trabasso and P. Van Den Broek, "Causal thinking and the representation of narrative events," *J. Memory Lang.*, vol. 24, no. 5, pp. 612–630, 1985.

[30] B. J. Grosz, A. Joshi, and S. Weinstein, "Centering: A framework for modeling the local coherence of discourse," *Comput. Linguistics*, vol. 21, no. 2, pp. 203–225, 1995.

[31] B. Wellner, J. Pustejovsky, C. Havasi, A. Rumshisky, and R. Sauri, "Classification of discourse coherence relations: An exploratory study using multiple knowledge sources," in *Proc. 7th SIGDIAL Workshop Discourse Dialogue*, 2006, pp. 117–125.

[32] G. Mulder and T. J. Sanders, "Causal coherence relations and levels of discourse representation," *Discourse Processes*, vol. 49, no. 6, pp. 501–522, 2012.

[33] J. R. Anderson, *Cognitive Psychology and its Implications*. New York, NY, USA: Macmillan, 2005.

[34] M. Dragoni, I. Donadello, and E. Cambria, "OntoSenticNet 2: Enhancing reasoning within sentiment analysis," *IEEE Intell. Syst.*, vol. 37, no. 2, pp. 103–110, Mar./Apr. 2022.

[35] A. Esuli, A. Moreo, and F. Sebastiani, "Cross-lingual sentiment quantification," *IEEE Intell. Syst.*, vol. 35, no. 3, pp. 106–114, May/Jun. 2020.

[36] M. Thelwall, "This! identifying new sentiment slang through orthographic pleonasm online: Yasss slay gorg queen ilysm," *IEEE Intell. Syst.*, vol. 36, no. 4, pp. 114–120, Jul./Aug. 2021.

[37] W. Li, W. Shao, S. Ji, and E. Cambria, "Bieru: Bidirectional emotional recurrent unit for conversational sentiment analysis," *Neurocomputing*, vol. 467, pp. 73–82, 2022.

[38] S. Modi and M. H. Bohara, "Facial emotion recognition using convolution neural network," in *Proc. 5th Int. Conf. Intell. Comput. Control Syst.*, 2021, pp. 1339–1344.

[39] S. Zhao et al., "A two-stage 3D CNN based learning method for spontaneous micro-expression recognition," *Neurocomputing*, vol. 448, pp. 276–289, 2021.

[40] W. Peng, X. Hong, and G. Zhao, "Adaptive modality distillation for separable multimodal sentiment analysis," *IEEE Intell. Syst.*, vol. 36, no. 3, pp. 82–89, May/Jun. 2021.

[41] N. Xu, W. Mao, P. Wei, and D. Zeng, "MDA: Multimodal data augmentation framework for boosting performance on sentiment/emotion classification tasks," *IEEE Intell. Syst.*, vol. 36, no. 6, pp. 3–12, Nov./Dec. 2021.

[42] A. Zadeh, R. Zellers, E. Pincus, and L.-P. Morency, "Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages," *IEEE Intell. Syst.*, vol. 31, no. 6, pp. 82–88, Nov./Dec. 2016.

[43] K. Zhang, Y. Li, J. Wang, E. Cambria, and X. Li, "Real-time video emotion recognition based on reinforcement learning and domain knowledge," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1034–1047, Mar. 2022.

[44] Y. M. Lee, Y. Chen, S. Li, and C.-R. Huang, "Emotion cause events: Corpus construction and analysis," in *Proc. 7th Int. Conf. Lang. Resour. Eval.*, 2010, pp. 1121–1128.

[45] A. Neviarouskaya and M. Aono, "Extracting causes of emotions from text," in *Proc. 6th Int. Joint Conf. Natural Lang. Process.*, 2013, pp. 932–936.

[46] I. Russo, T. Caselli, F. Rubino, E. Boldrini, and P. Mart ínez-Barco, "Emocause: An easy-adaptable approach to extract emotion cause contexts," in *Proc. 2nd Workshop Comput. Approaches Subjectivity Sentiment Anal.*, 2011, pp. 153–160.

[47] K. Gao, H. Xu, and J. Wang, "Emotion cause detection for chinese micro-blogs based on ecocc model," in *Proc. Pacific-Asia Conf. Knowl. Discov. Data Mining*, 2015, pp. 3–14.

[48] L. Gui, D. Wu, R. Xu, Q. Lu, and Y. Zhou, "Event-driven emotion cause extraction with corpus construction," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2016, pp. 1639–1649.

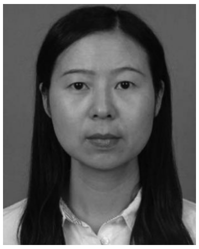
[49] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[50] X. Li, K. Song, S. Feng, D. Wang, and Y. Zhang, "A co-attention neural network model for emotion cause analysis with emotional context awareness," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2018, pp. 4752–4757.

[51] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[52] L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, and J. Du, "A question answering approach for emotion cause extraction," in *Proc. 2017 Conf. Empir. Methods Natural Lang. Process.*, 2017, pp. 1593–1602.

[53] B. Xu, H. Lin, Y. Lin, Y. Diao, L. Yang, and K. Xu, "Extracting emotion causes using learning to rank methods from an information retrieval perspective," *IEEE Access*, vol. 7, pp. 15 573–15 583, 2019.
[54] G. Hu, G. Lu, and Y. Zhao, "Bidirectional hierarchical attention networks based on document-level context for emotion cause extraction," in *Proc. Findings Assoc. Comput. Linguistics*, 2021, pp. 558–568.
[55] R. Xia and Z. Ding, "Emotion-cause pair extraction: A new task to emotion analysis in texts," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 1003–1012.
[56] Z. Ding, R. Xia, and J. Yu, "Ecpe-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3161–3170.
[57] X. Chen, Q. Li, and J. Wang, "A unified sequence labeling model for emotion cause pair extraction," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 208–218.
[58] C. Fan, C. Yuan, J. Du, L. Gui, M. Yang, and R. Xu, "Transition-based directed graph construction for emotion-cause pair extraction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3707–3717.
[59] P. Wei, J. Zhao, and W. Mao, "Effective inter-clause modeling for end-to-end emotion-cause pair extraction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3171–3181.
[60] Y. Chen, W. Hou, X. Cheng, and S. Li, "Joint learning for emotion classification and emotion cause detection," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2018, pp. 646–651.
[61] J. W. Orr, "Towards narrative understanding with deep neural networks and hidden Markov models," Ph.D. dissertation, Oregon State University, Corvallis, OR, USA, 2019.
[62] S. Benini, M. Savardi, K. Bálint, A. B. Kovács, and A. Signoroni, "On the influence of shot scale on film mood and narrative engagement in film viewers," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 592–603, Apr./Jun. 2022.
[63] J. Tarvainen, J. Laaksonen, and T. Takala, "Film mood and its quantitative determinants in different types of scenes," *IEEE Trans. Affect. Comput.*, vol. 11, no. 2, pp. 313–326, Apr./Jun. 2020.
[64] J. Tang et al., "From discourse to narrative: Knowledge projection for event relation extraction," 2021, *arXiv:2106.08629*.
[65] J. Tourille, O. Ferret, A. Neveol, and X. Tannier, "Neural architecture for temporal relation extraction: A Bi-LSTM approach for detecting narrative containers," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 224–230.
[66] A. Gupta, H. Abi-Akl, and H. De Mazancourt, "Not all titles are created equal: Financial document structure extraction shared task," in *Proc. 3rd Financial Narrative Process. Workshop*, 2021, pp. 86–88.
[67] T. Kočiskỳ et al., "The narrativeqa reading comprehension challenge," *Trans. Assoc. Comput. Linguistics*, vol. 6, pp. 317–328, 2018.
[68] J. Fitzgerald and D. L. Spiegel, "Enhancing children's reading comprehension through instruction in narrative structure," *J. Reading Behav.*, vol. 15, no. 2, pp. 1–17, 1983.
[69] S. Babayiğit, S. Roulstone, and Y. Wren, "Linguistic comprehension and narrative skills predict reading ability: A 9-year longitudinal study," *Brit. J. Educ. Psychol.*, vol. 91, no. 1, pp. 148–168, 2021.
[70] R. C. Schank and R. P. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures*, London, U.K.: Psychology Press, 2013.
[71] X. Zhou, S. Luo, and Y. Wu, "Co-attention hierarchical network: Generating coherent long distractors for reading comprehension," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 9725–9732.
[72] A. Bolte, T. Goschke, and J. Kuhl, "Emotion and intuition: Effects of positive and negative mood on implicit judgments of semantic coherence," *Psychol. Sci.*, vol. 14, no. 5, pp. 416–421, 2003.
[73] J. Chen, J. Chen, and Z. Yu, "Incorporating structured commonsense knowledge in story completion," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 6244–6251.
[74] D. Herremans and E. Chew, "Morpheus: Generating structured music with constrained patterns and tension," *IEEE Trans. Affect. Comput.*, vol. 10, no. 4, pp. 510–523, Oct./Dec. 2019.
[75] S. Cumming, G. Greenberg, and R. Kelly, "Conventions of viewpoint coherence in film," *Philosophers*, vol. 17, no. 1, pp. 1–29, 2017.
[76] Z. Hu and M. Walker, "Inferring narrative causality between event pairs in films," in *Proc. 18th Annu. SIGDIAL Meeting Discourse Dialogue*, 2017, pp. 342–351.
[77] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol.*, 2019, pp. 4171–4186.
[78] Y. Cui et al., "Pre-training with whole word masking for chinese bert," 2019, *arXiv:1906.08101*.
[79] J. Langer, "The reading process," *Secondary School Reading: What Research Reveals for Classroom Practice*, Berlin, Germany: Springer, 1982, pp. 39–52.
[80] J. L. Myers, M. Shinjo, and S. A. Duffy, "Degree of causal relatedness and memory," *J. Memory Lang.*, vol. 26, no. 4, pp. 453–465, 1987.
[81] T. Trabasso and M. Langston, "Modeling causal integration and availability of information during comprehension of narrative texts," *Construction Ment. Representations During Reading*, vol. 25, pp. 25–59, 1998.
[82] P. van Den Broek, B. Linzie, C. Fletcher, and C. J. Marsolek, "The role of causal discourse structure in narrative writing," *Memory Cogn.*, vol. 28, no. 5, pp. 711–721, 2000.
[83] R. Xu, J. Hu, Q. Lu, D. Wu, and L. Gui, "An ensemble approach for emotion cause detection with event extraction and multi-kernel SVMs," *Tsinghua Sci. Technol.*, vol. 22, no. 6, pp. 646–659, 2017.
[84] G. Montavon, G. Orr, and K.-R. Müller, *Neural Networks: Tricks of the Trade*, Berlin, Germany: Springer, 2012, vol. 7700.
[85] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Informat. Process. Syst.*, 2013, pp. 3111–3119.
[86] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*.
[87] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Their Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998.
[88] K. Bardovi-Harlig, "Reverse-order reports and the acquisition of tense: Beyond the principle of chronological order," *Lang. Learn.*, vol. 44, no. 2, pp. 243–282, 1994.
[89] L. Xu and H. Lin, "Ontology-driven affective chinese text analysis and evaluation method," in *Proc. Int. Conf. Affect. Comput. Intell. Interaction*, 2007, pp. 723–724.



Wei Cao is currently working toward the PhD degree in data mining with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China. Her current research interests include affective computing, computer vision, natural language processing and text mining.

Kun Zhang received the PhD degree in computer science and technology from the University of Science and Technology of China, Hefei, China, in 2019. He is is currently a faculty member with the Hefei University of Technology (HFUT), China. His research interests include natural language processing, Recommendation System, and text mining. He has published several papers in refereed conference proceedings such as AAAI, KDD, ICDM. He received the KDD 2018 Best Student Paper Award.

Shulan Ruan received the BS degree from Hunan University, Changsha, China, in 2018. He is currently working toward the PhD degree with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China. His research interests include sentiment analysis, computer vision and natural language processing. He has published several papers in *IEEE Transactions on Multimedia*, ICCV, AAAI, ICME, BIBM, etc.

Hanqing Tao is currently working toward the PhD degree with the Department of Computer Science and Technology from University of Science and Technology of China (USTC). His research interests include data mining, natural language processing, Chinese language analysis and interpretable artificial intelligence. He has published several papers in referred conference proceedings, such as AAAI, ICDM, ICME, CCL etc.

Sirui Zhao is currently working toward the PhD degree with the Department of Computer Science and Technology from University of Science and Technology of China (USTC). His research interests include automatic micro-expressions analysis, human-computer interaction (HCI) and affect computing. He has published several papers in refereed conferences and journals, including *ACM Multimedia Conference*, Neural Networks, *Neurocomputing*, etc.

Hao Wang received the PhD degree in computer science from USTC. He is currently an associate researcher with the School of Computer Science and Technology, USTC. His main research interests include data mining, representation learning, network embedding and recommender systems. He has published several papers in referred conference proceedings, such as TKDE, TOIS, NeuriPS, and AAAI.

Qi Liu (Member, IEEE) received the PhD degree in computer science from USTC. He is a professor with USTC. His general area of research is datamining and knowledge discovery. He has published prolifically in refereed journals and conference proceedings, e.g., *IEEE Transactions on Knowledge and Data Engineering*, *ACM Transactions on Information Systems*, *ACM Transactions on Knowledge Discovery from Data*, *ACM Transactions on Intelligent Systems and Technology*, KDD, IJCAI, AAAI, ICDM, SDM, and CIKM. He is a member of the ACM. He received the ICDM 2011 Best Research Paper Award and the Best of SDM 2015 Award.

Enhong Chen (Senior Member, IEEE) received the PhD degree from USTC. He is a professor and vice dean with the School of Computer Science, USTC. His general area of research includes data mining and machine learning, social network analysis, and recommender systems. He has published more than 100 papers in refereed conferences and journals, including *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Mobile Computing*, KDD, ICDM, NeurIPS, and CIKM. He was on program committees of numerous conferences including KDD, ICDM, and SDM. His research is supported by the National Science Foundation for Distinguished Young Scholars of China.