Improving Affective Event Classification with Multi-Perspective Knowledge Injection

Wenjia Yi, Yanyan Zhao*, Jianhua Yuan, Weixiang Zhao, Bing Qin
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{wjyi, yyzhao, jhyuan, wxzhao, qinb}@ir.hit.edu.cn

Abstract

In recent years, many researchers have recognized the importance of associating events with sentiments. Previous approaches focus on generalizing events and extracting sentimental information from a large-scale corpus. However, since context is absent and sentiment is often implicit in the event, these methods are limited in comprehending the semantics of the event and capturing effective sentimental clues. In this work, we propose a novel Multi-perspective Knowledge-injected Interaction Network (MKIN) to fully understand the event and accurately predict its sentiment by injecting multi-perspective knowledge. Specifically, we leverage contexts to provide sufficient semantic information and perform context modeling to capture the semantic relationships between events and contexts. Moreover, we also introduce human emotional feedback and sentiment-related concepts to provide explicit sentimental clues from the perspective of human emotional state and word meaning, filling the reasoning gap in the sentiment prediction process. Experimental results on the gold standard dataset show that our model achieves better performance over the baseline models.

1 Introduction

Affective Event Classification (AEC) aims at predicting the sentiment polarity of the given event. We consider events that have positive effects on people who experience them as positive events. For instance, typically positive events include *having a new crush*, *going to the bonfire*, *seeing a rainbow*. On the contrary, events that have negative effects on people who experience them are treated as negative events, such as *breaking a marriage*, *going to the funeral*, *hearing a loud noise*. Since events often trigger sentiments and sentiments are often implicit, recognizing affective events is of great values to various natural language processing applications, covering dialogue systems (Shi and Yu, 2018), question-answering systems (Oh et al., 2012), implicit sentiment analysis (Zhou et al., 2021) and opinion mining (Xu et al., 2022b).

The challenges of AEC lie in the limited context and the implicit sentiment of the event. To be specific, we often rely on the context to analyze sentiments, but there is no rich context to understand the event. Besides, traditional sentiment analysis methods rely on the occurrence of explicit sentiment words, but there are few explicit sentimental clues in the event. Many previous approaches have been devoted to cope with the challenges by extracting sentimental information from a large-scale corpus (Ding and Riloff, 2018; Saito et al., 2019; Zhuang et al., 2020). However, such attempts are not effective enough to understand the event and capture sentimental clues due to weak context modeling and insufficient sentimental information.

We believe that this task would benefit from multi-perspective knowledge injection. Specifically, context can provide additional semantic information to understand the event, and human emotional feedback as well as sentiment-related concepts can provide explicit sentimental information from two perspectives to fill the reasoning gap. Figure 1 shows an example of an affective event, demonstrating the significance

*Corresponding author

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

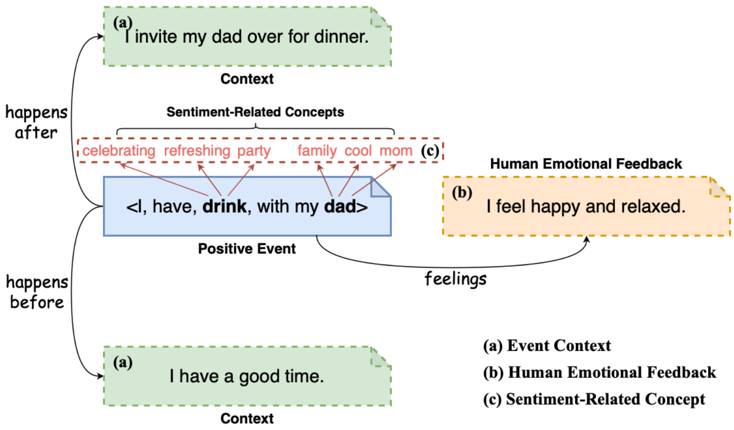

Figure 1: An example of an affective event for identifying the sentiment with the help of event contexts, human emotional feedback and sentiment-related concepts.

of contexts, human emotional feedback and sentiment-related concepts in understanding events and detecting implied sentiments. On the one hand, the given event ("⟨I, have, drink, with my dad⟩") could be thoroughly understood by supplying contexts ("I invite my dad over for dinner.", "I have a good time."). On the other hand, from people's positive emotional feedback towards the event ("happy", "relaxed"), we could know that the event would typically have a positive effect on them. Moreover, the meaning of "drink" and "dad" in the event are enriched by sentiment-related concepts ("celebrating", "party", "cool", etc.). Thus, the implicit sentiment of the event could be identified more easily via enhanced emotions and enriched concepts.

In this paper, to cope with the aforementioned challenges, we propose a novel Multi-perspective Knowledge-injected Interaction Network (MKIN) to thoroughly comprehend the event and precisely infer its sentiment. Specifically, we utilize a pre-trained generative commonsense reasoning model to create event contexts and human emotional feedback of the event. Meanwhile, a commonsense knowledge base and an emotion dictionary are adopted to retrieve sentiment-related concepts. Then, we devise a Multi-Source Text Encoding Module to encode these events and knowledge. To better integrate contextual information and sentimental clues, we construct a Semantic and Sentimental Fusion Module, which performs interaction as well as fusion of semantics and sentiments. Finally, we introduce a classifier to accurately classify affective events.

To evaluate the performance of MKIN, we conduct extensive experiments on the gold standard dataset for AEC. State-of-the-art performance is achieved by us compared with the baseline models.

The main contributions of our work are summarized as follows:

- For the first time, we propose to leverage commonsense knowledge to improve Affective Event Classification.
- We introduce a novel approach MKIN to perform context modeling and sentiment reasoning, which injects knowledge from multiple perspectives to meet the challenges of AEC.
- Extensive experimental results on the benchmark dataset demonstrate the superiority of MKIN. Our source code will be publicly available.

2 Related Work

Relevant work mainly includes two directions, one is affective event classification, and the other is incorporating external knowledge in sentiment analysis tasks.

2.1 Affective Event Classification

Prior work has focused on producing lexical resources of verbs or event phrases with corresponding sentiment polarity values. Goyal et al. (2010) created a new type of lexicon for narrative text comprehen-

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.

sion, consisting of patient polarity verbs that impart positive or negative states on their patients. Vu et al. (2014) created a manually-constructed dictionary of emotion-provoking events, then used seed expansion and clustering to automatically acquire and aggregate events from web data. Li et al. (2014) extracted major life events from Twitter by clustering tweets corresponding to speech act words, such as ”congratulations” or ”condolences”. Ding and Riloff (2016) first defined stereotypical affective events as triples $\langle$Agent, Verb, Object$\rangle$ that are independent of context, and used a semi-supervised label propagation algorithm to discover affective events from Blogs.

More recently, many researches on affective event classification exert much effort to extract sentimental information from a large-scale corpus. Ding and Riloff (2018) expanded affective events as tuples $\langle$Agent, Predicate, Theme, Prepositional Phrase$\rangle$, and introduced a weakly supervised semantic consistency model for inducing a large collection of affective events from a personal story corpus. Saito et al. (2019) proposed to exploit discourse relations to propagate sentiment polarity from seed predicates. They extracted events that co-occur with seeds in a Japanese web corpus, and used discourse relations as constraints in the model learning process. Zhuang et al. (2020) first utilized the pre-trained model and presented a discourse-enhanced self-training method, which combines the classifier’s predictions with information from local discourse contexts, and iteratively improves the classifier with unlabeled data. Another line of related work is Event-related Sentiment Analysis, which explicitly models events to improves sentiment analysis because events often trigger sentiments in sentences. Zhou et al. (2021) proposed a hierarchical tensor-based composition mechanism for event-centered text representation and develop a multi-task learning framework to improve sentiment analysis with event type classification.

However, existing methods only induce affective events based on semantic relations or discourse relations, or purely focus on sentimental information from local discourse contexts. Unlike the previous work, we consider multiple perspectives of knowledge, covering event contexts, human emotional feedback and sentiment-related concepts.

2.2 Sentiment Analysis with Knowledge

In recent years, there is a growing number of researches on incorporating external knowledge in various sentiment analysis tasks. Turcan et al. (2021) explored the use of commonsense knowledge via adapted knowledge models to understand implicitly expressed emotions and the reasons of those emotions for Emotion Cause Extraction. Sabour et al. (2021) leveraged commonsense knowledge to obtain more information about the user’s situation and feelings to further enhance the empathy expression in the generated responses for Empathetic Response Generation. Zhao et al. (2022) utilized commonsense knowledge to provide causal clues to guide the process of causal utterance traceback for Emotion Recognition in Conversations. Peng et al. (2022) employed commonsense knowledge to obtain the psychological intention of the help-seeker to generate the supportive responses for Emotional Support Conversation. Xu et al. (2022a) used a knowledge graph to supplement a large amount of knowledge and common sense omitted in implicit emotional sentences for Implicit Sentiment Analysis.

There are also many studies on integrating external knowledge in other natural language processing tasks, but less studies on Affective Event Classification. To the best of our knowledge, this is the first attempt to introduce external knowledge into Affective Event Classification task.

3 Methodology

The problem of the AEC task could be formulated as follows. Given an event tuple $\langle$ Agent: $agent = \{w_1, w_2, \cdots, w_{n_{agent}}\}$, Predicate: $pred = \{w_1, w_2, \cdots, w_{n_{pred}}\}$, Theme: $theme = \{w_1, w_2, \cdots, w_{n_{theme}}\}$, Prepositional Phrase: $prep = \{w_1, w_2, \cdots, w_{n_{prep}}\}$ $\rangle$ with the corresponding sentiment category, the goal of this task is to predict the sentiment distribution over three sentiment polarities.

The overall architecture of our proposed model MKIN is shown in Figure 2, which consists of four modules: Knowledge Acquisition Module, Multi-Source Text Encoding Module, Semantic and Sentimental Fusion Module, Sentiment Classification Module. Each one of the four modules will be elaborated in the rest of this section.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

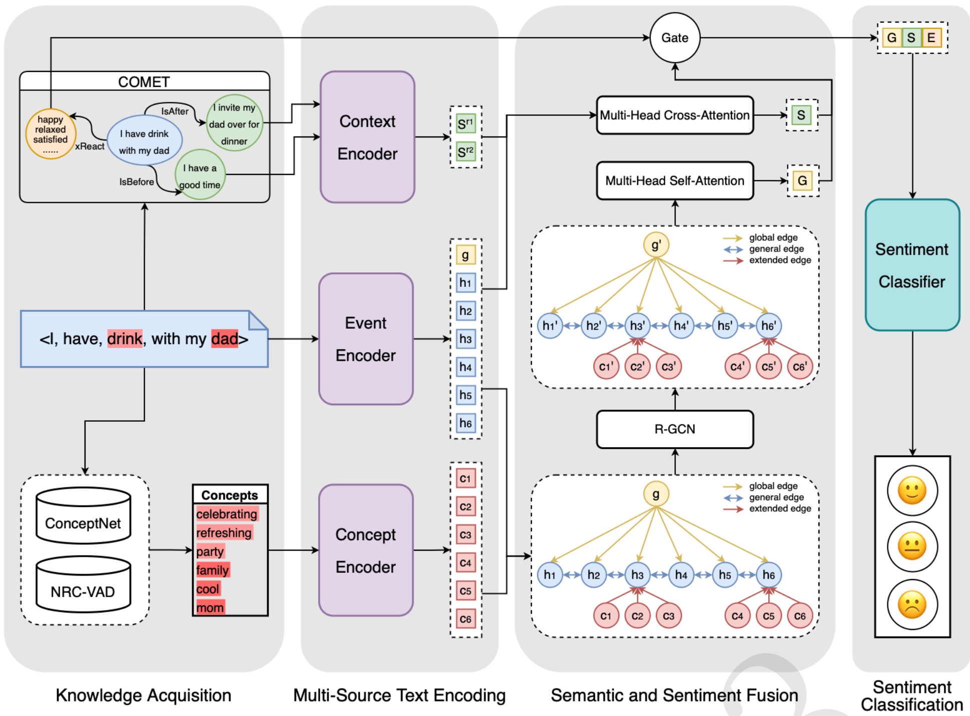

Figure 2: The overall architecture of our proposed model.

3.1 Knowledge Acquisition Module

Event Context Acquisition. Since retrieving context from corpus is expensive and noisy, we turn to the commonsense knowledge base to provide the context for the given event. In this work, we employ ATOMIC-2020 (Hwang et al., 2021) as our commonsense knowledge base, which is a commonsense knowledge graph of general-purpose everyday inferential knowledge covering social, physical, and event-centered aspects.

To be more specific, we explore two event-centered categories of commonsense knowledge from ATOMIC-2020, called *isAfter* and *isBefore*. These two relations provide reasoning about event scripts or sequences, respectively introducing events that can precede or follow an event. Therefore, we use *isAfter* and *isBefore* to introduce two events that happened before and happened after the given event, respectively. The two introduced events could form the context of a given event, and the three events can be treated as a narrative event chain. Then the given event could be fully understood via context awareness.

In order to acquire contexts for given events, we adopt a generative model COMET (Bosselut et al., 2019) which is a pre-trained GPT-2 model (Radford et al., 2018) finetuned on ATOMIC (Sap et al., 2019). More precisely, we use a BART-based (Lewis et al., 2020) variation of COMET, which is trained on ATOMIC-2020. This model can generate accurate and representative knowledge for new, unseen events. It is suitable and necessary for AEC task, because affective event has a broad scope and many events may not exist in the static ATOMIC-2020 dataset. An event is given to form the input format $(e, r, [GEN])$, where e is the sequence that comprise an event tuple. For instance, $\langle$I, have, drink, with my dad$\rangle$ is converted into the sequence "I have drink with my dad". And r is the relations we select, including *isAfter* and *isBefore*. Then we use COMET to generate five commonsense inferences for each relation r.

Human Emotional Feedback Acquisition. In this work, ATOMIC-2020 is also utilized to acquire human emotional feedback. We explore one type of social-interaction commonsense knowledge called *xReact*, which manifests the emotional states of the participants in a given event. The introduced emotion reactions could fill the reasoning gap between events and sentiments. We acquire human emotional

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

Dimensions	Values	Interpretations
Valence	[0,1]	Negative - Positive
Arousal	[0,1]	Calm - Excited
Dominance	[0,1]	Submissive - Dominant

Table 1: Interpretations of NRC_VAD dimensions

feedback in the same way that we acquire event contexts. As the commonsense inferences for *xReact* are usually emotion words (e.g., happy, sad, angry, etc.) rather than events or sentences, we simply adopt the hidden state representation from the last encoder layer of COMET as the human emotional feedback representation.

Sentiment-Related Concept Acquisition. Following (Li et al., 2022b), we use a commonsense knowledge base ConceptNet (Speer et al., 2017) combined with an emotion lexicon NRC_VAD (Mohammad, 2018) to obtain sentiment-related concepts.

ConceptNet is a large-scale multilingual semantic graph proposed to describe general human knowledge, allowing natural language applications to better understand the meanings behind the words. We introduce the tuple *(head concept, relation, tail concept, confidence score)* to represent the assertions in ConceptNet graph and their associated confidence scores, and denote the tuple as $\tau = (h, r, t, c)$. For instance, one such tuple from Conceptnet is *(birthday, RelatedTo, happy, 4.16)*. Let W be a collection of words in a given event tuple. For each non-stopword $h \in W$, we retrieve a set of tuples $T_i = \left\{\tau_i^j = \left(h_i, r_i^j, t_i^j, c_i^j\right)\right\}$ containing its immediate neighbors from ConceptNet, where i, j are indices of non-stopwords and the retrieved tuples.

To refine the retrieved concepts, we first remove tuples where concepts t_i^j are stopwords or not in our vocabulary. We further filter tuples where confidence scores c_i^j are smaller than 1 to reduce annotation noises. As many of the tuples are still useless for our AEC task, we select 10 relevant relations from 38 relations in ConceptNet as (Liao et al., 2022) did, they analyzed the effects of various relations on implicit sentiment analysis in detail. And then we remove the tuples where relations r_i^j belong to other relations.

To highlight sentimental information, we adopt NRC_VAD to measure sentimental intensity of the external concepts. NRC_VAD is a lexicon with valence, arousal, and dominance (VAD) scores. The interpretations of three dimensions are presented in Table 1. Such as the VAD score vector $[V_a, A_r, D_o]$ of word "happy" is $[1.000, 0.735, 0.772]$. Following (Zhong et al., 2019), sentimental intensity value of a concept x is computed as:

$$\eta(x) = min\text{-}max\left(\left\|V_a(x) - \frac{1}{2}, \frac{A_r(x)}{2}\right\|_2\right) \tag{1}$$

where $min\text{-}max()$ denotes min-max normalization, $\|.\|_k$ denotes L_k norm, $V_a(x)$ and $A_r(x)$ denote the valence and arousal scores in VAD vector of concept x, respectively. For concept x not in NRC_VAD, $\eta(x)$ will be set to 0. We rank the tuples according to the sentimental intensity values $\eta(t_i^j)$ of concepts t_i^j. Based on the order of tuples, we reserve at most three external concepts with adequate sentimental intensity values (i.e., $\eta(t_i^j) \geq 0.6$) for each word h.

3.2 Multi-Source Text Encoding Module

Multi-source text encoder considers text from three sources, including raw event, event context and external concepts. The event encoder, context encoder and concept encoder are the same encoder, which employ widely-used pre-trained model BERT (Devlin et al., 2019).

Firstly, the events are encoded. For each event $e = \{w_1, w_2, \cdots, w_n\}$, we concatenate two special tokens $[CLS]$ and $[SEP]$ to the beginning and end of the event. Then the sequence $\{[CLS], w_1, w_2, \cdots, w_n, [SEP]\}$ is fed to the encoder, leading to a series of hidden states:

$$h_i = \text{BERT}(\,[CLS], w_1, w_2, \cdots, w_n, [SEP]\,) \tag{2}$$

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.

where $h_i \in \mathbb{R}^{d_m}$ is the i-th token in the input sequence, d_m is the dimension of hidden states in BERT. And the vectorized representation of an event is H. It is worth noting that we specifically denote $[CLS]$ token as g.

Secondly, the contexts are encoded. For both *isAfter* and *isBefore* , we concatenate the five generated commonsense inferences to get a context sequence CS^r:

$$CS^r = cs_1^r \oplus cs_2^r \cdots \oplus cs_5^r \tag{3}$$

where $\oplus$ is the concatenation operation, r is the relation we select from ATOMIC-2020. Then we pass each context sequence in the same input format as $\{[CLS], CS_r, [SEP]\}$, to derive a series of hidden states from the last layer:

$$s_j^r = \text{BERT}(\,[CLS], CS_r, [SEP]\,) \tag{4}$$

where $s_j^r \in \mathbb{R}^{d_m}$ is the j-th token in the input sequence. And the vectorized representation of a context sequence is S^r.

Thirdly, the concepts are encoded. For each concept x, we perform mean-pooling operation from the last hidden layer to obtain its representation $c \in \mathbb{R}^{d_m}$:

$$c = \text{Mean-pooling}(\,\text{BERT}(\,[CLS], x, [SEP]\,)\,) \tag{5}$$

3.3 Semantic and Sentimental Fusion Module

As contexts and sentimental clues have been collected, Semantic and Sentimental Fusion Module is devised to perform interaction as well as fusion of contextual and sentimental information.

Semantic Interaction. In order to highlight the more important semantic features from the contexts, we utilize multi-head cross-attention mechanism (Vaswani et al., 2017) to achieve the interaction of contexts and the event. Then for each context sequence CS^r, a context-aware representation $S^{r\prime}$ is learned:

$$S^{r\prime} = \text{MH}(f(H), f(S^r), f(S^r)) \tag{6}$$

where f is a linear transformation, each vector is transformed to the dimension of d_h with f, and

$$\text{MH}(Q, K, V) = \text{Concat}(head_1, \cdots, head_h)W^O \tag{7}$$

$$head_i = \text{Att}(QW_i^Q, KW_i^K, VW_i^V) \tag{8}$$

$$\text{Att}(Q, K, V) = \text{Softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{9}$$

where Q, K, and V are sets of queries, keys and values, respectively, the projections are parameter matrices $W^O \in \mathbb{R}^{md_v \times d_h}$, $W_i^Q \in \mathbb{R}^{d_h \times d_k}$, $W_i^K \in \mathbb{R}^{d_h \times d_k}$, $W_i^V \in \mathbb{R}^{d_h \times d_v}$, and $d_k = d_v = d_h/h$. The final context representation S is obtained by:

$$S = \bigoplus_{r \in \{isAfter, isBefore\}} \text{Max-pooling}(S^{r\prime}) \tag{10}$$

then S is transformed to the dimension of d_h with a linear projection.

Sentimental Interaction. We construct a graph network for modeling the event and relevant concepts. Specifically, each event token and concept are represented as vertices in the graph, including the $[CLS]$ token as a global vertex for aggregating information. Furthermore, three relation types of edges are applied to connect the vertices: (1) *global edge*, a directed edge which connects global node to each event node; (2) *general edge*, an undirected edge between two successive event nodes; (3) *extended edge*, a directed edge which connects a concept node to the corresponding event node.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ denotes our graph, where $\mathcal{V}$, $\mathcal{E}$, and $\mathcal{R}$ are sets of vertices, edges and relation types, respectively. We initialize each vertex with the corresponding encoded feature vector, and denote vertex features as $V = \{g, h_1, \cdots, h_n, c_1, \cdots, c_m\} = \{v_1, v_2, \cdots, v_N\}$.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.

We feed the initial vertex features into a graph encoder to propagate semantic and sentimental information. Considering different relation types of edges, we adopt relational graph convolutional networks (Schlichtkrull et al., 2018) to update vertex representations. The convolutional computation for a vertex at the $(l+1)$-th layer which takes the representation $v_i^{(l)}$ at the l-th layer as input is defined as:

$$v_i^{(l+1)} = \text{ReLU}\left(\sum_{r\in\mathcal{R}}\sum_{v\in\mathbb{N}_i^r}\frac{1}{|\mathbb{N}_i^r|}W_r^{(l)}v_i^{(l)}\right) \tag{11}$$

where ReLU (Agarap, 2018) is an activation function, $\mathbb{N}_i^r$ is the set of neighbor vertices under relation type r, and $W_r^{(l)}$ are relation-specific learnable parameters at the l-th layer.

To selectively attend to the more important sentimental features within the enriched event representation, we pass the updated vertex features to a multi-head self-attention layer, then a sentiment-enhanced representation V' is learned:

$$V' = \text{MH}(V^{(L)}, V^{(L)}, V^{(L)}) \tag{12}$$

where $V^{(L)} = \left\{v_1^{(L)}, v_2^{(L)}, \cdots, v_N^{(L)}\right\} = \{g', h'_1, \cdots, h'_n, c'_1, \cdots, c'_m\}$ are the outputs of the graph encoder. We take the global vertex feature as the final event representation G, and G is transformed to the dimension of d_h with a linear projection.

Feature Fusion. We first transform the human emotional feedback representation E to the dimension of d_h via a linear transformation. Inspired by (Liu et al., 2021), we fuse the three representations with a gated manner, including context representation S, event representation G, and human emotional feedback representation E. The gate is formulated as:

$$G_S = \text{ReLU}(\text{FC}([G, S, G - S, G \odot S])) \tag{13}$$

$$G_E = \text{ReLU}(\text{FC}([G, E, G - E, G \odot E])) \tag{14}$$

$$p = \text{Sigmoid}(\text{FC}[G_S, G_E]) \tag{15}$$

where FC is a fully-connected layer and $[.,.]$ means concatenation. Then the three features are fused as:

$$F = G + p \odot S + (1 - p) \odot E \tag{16}$$

3.4 Sentiment Classification Module

Finally, taking the above fused representation as input, a sentiment classifier is applied to predict the sentiment of the event:

$$\hat{y} = \text{Softmax}(\text{MLP}(F)) \tag{17}$$

where MLP is a multi-layer perception.

Cross entropy loss is adopted to train the model, the loss function is defined as:

$$\mathcal{L} = -\frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{C} y_i^j \cdot log(\hat{y}_i{}^j) \tag{18}$$

where T and C denote the number of training examples and the number of sentiment categories, respectively, and y_i^j represents the ground-truth label.

4 Experiments

In this section we present the dataset, evaluation metrics, baseline models, model variants, and other experimental settings.

Category	Number
Negative Event	348
Neutral Event	717
Positive Event	435

Table 2: Dataset statistics

4.1 Dataset and Evaluation Metrics

We conduct experiments on the gold standard dataset for AEC. It is collected from Twitter Dataset with sentiment category labels annotated by (Zhuang et al., 2020), and the sentiment categories belong to negative, neutral and positive. Statistics of the dataset are shown in Table 2.

Following (Zhuang et al., 2020), we report the precision, recall and F1 score for each of the three categories, and weighted average results for each metric.

4.2 Baselines and Comparison Models

We compare our proposed model with the following method:

BERT-base/large (Devlin et al., 2019): BERT is a widely-used pre-trained language model with excellent performance in various natural language processing tasks. We adopt the base version and the large version of BERT as the basis for our classifier and perform fine-tuning during the training process.

RoBERTa-base/large (Liu et al., 2019): RoBERTa has the same model architecture as BERT but with a robustly optimized pre-training scheme allowing it to generalize better to downstream tasks. Similarly, we adopt the base version and the large version of RoBERTa for experiments.

DEST (Zhuang et al., 2020): DEST is a discourse-enhanced self-training model which is the state-of-the-art model for AEC. It introduces BERT-base model for classification and combines the classifier's predictions with information from local discourse contexts to iteratively assign high-quality labels to new training instances.

4.3 Implementation Details

Following (Zhuang et al., 2020), we performed 10-fold cross-validation over the dataset, where each of the 10 runs used 8 folds of the data for training, 1 fold of the data for validation and tuning, and 1 fold of the data for testing.

Base version of BERT is adopted as the encoder, and the dimension of hidden states d_m in the encoder is 768. For all representations in the rest of our model, the dimension d_h is set to 300. For the multi-head cross-attention layer and the multi-head self-attention layer, the number of attention head is 5 and 12, respectively. For sentiment classification, the dimensions of MLP are set to [300, 100, 3] and the dropout rate is set to 0.1. We train our model with AdamW optimizer in a learning rate of 1e-5 and a linear warmup rate of 0.1. And the batch size is set to 8. We implemented all models in PyTorch with a single Tesla V100 GPU. Reported results are medians over 5 times of 10-fold cross-validation with the same 5 distinct random seeds.

5 Results and Analysis

In this section we present model evaluation results, ablation study, and case study.

5.1 Overall Results

As depicted in Table 3, our proposed model achieves state-of-the-art results. Benefiting from the effective context modeling with event contexts and accurate sentiment reasoning with human emotional feedback and sentiment-related concepts, MKIN achieves the best results on each metrics and the highest F1 score in each category compared with the state-of-the-art model DEST and other baselines.

For the state-of-the-art model DEST, we reproduce the performance in the same setting as the original model. Although DEST utilizes a large number of coreferent sentiment expressions to provide explicit sentiment clues, it is unreliable because coreferent sentiment expressions are quite noisy due to imperfect

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.

Model	NEG			NEU			POS			P	R	F1
	P	R	F1	P	R	F1	P	R	F1			
BERT-base (110M)	71.6	77.4	74.1	76.5	78.1	77.1	76.8	69.9	73	75.8	75.3	75.3
BERT-large (340M)	72.5	75.5	73.5	76.7	78.6	77.5	77.6	72	74.3	76.4	75.7	75.7
RoBERTa-base (125M)	73.4	74.9	73.6	78	79.6	78.7	78.3	74.5	76	77.3	76.9	76.8
RoBERTa-large (355M)	74.4	75.3	74.5	77.7	82.3	79.8	78.9	71.7	74.8	77.7	77.3	77.2
DEST (110M)	78.9	**77.6**	78	78	83.7	80.6	**80.2**	71.5	75	79.2	78.6	78.5
MKIN (ours) (110M)	**83.7**	75.9	**79.1**	**80.1**	**84.3**	**82**	80	**78.8**	**79.1**	**81.3**	**80.7**	**80.6**

Table 3: Performance of all models. The best results among all models are highlighted in **bold**.

Model	P	R	F1
MKIN	**81.3**	**80.7**	**80.6**
w/o Event Context	79.9	79.3	79.2
w/o Human Emotional Feedback	79.7	79	78.9
w/o Sentiment-Related Concept	79.4	78.8	78.8
w/o R-GCN	79.8	78.9	78.8
w/o Gate	79.7	79.1	79.1

Table 4: Results of ablation study on model components.

coreference and issues like sarcasm, which leads to low-quality pseudo labels, even if an additional event classifier is introduced. Instead of retrieving information from corpus, we turn to the commonsense knowledge base for context information and explicit sentiment clues. MKIN improves precision of negative events from 78.9 to 83.7 and improves recall of positive events from 71.5 to 78.8. The substantial gain demonstrates the effectiveness of injecting multi-perspective knowledge to improve affective event classification, and shows the strong ability of our Semantic and Sentimental Fusion Module in extracting important features for enriching the event representation.

For other baselines models, they are not comparable with our proposed model MKIN. It suggests that the event representations extracted by pre-trained language models are not sufficient for classification, and only slight improvements are gained when a larger model is adopted. Besides, two instructive conclusions can be derived. On the one hand, it is of great significance to perform context modeling and capture semantic relationships between events and contexts, which lead to the thorough understanding of the events. On the other hand, explicit sentiment clues provided by human emotional feedback and sentiment-related concepts can fill the reasoning gap between events and sentiments.

5.2 Ablation Study

To gain better insight into the performance of our proposed model MKIN, we conduct an ablation study to verify the contributions of its main components.

Results in Table 4 show that each component is beneficial to the final performance. First, when the Event Context component is removed, the semantic features of the context are not integrated in the final representation of the event. The performance of the model degrades to a certain extent, which proves that context modeling is crucial to AEC. Since there is very limited information in the event, the model needs additional semantic information from the context for better event representation learning. Second, when removing the Human Emotional Feedback component, human's feelings are not taken into account. The dropped results demonstrate that human emotional feedback are powerful sentimental signals. Third, when the Sentiment-Related Concept component is removed, external concepts are not introduced to expand the original word meaning. The performance of the model decreased even more, which suggests that sentiment-related concepts have a considerable impact on sentiment classification. Introducing external sentimental commonsense knowledge and enriching the meaning of words in events can help the model detect implicit sentiments. Besides, the use of R-GCN enables more accurate capturing of interactive information, while the gate can better fuse complementary information.

Event & Label	Event Context	Human Emotional Feedback	Sentiment-Related Concept
⟨I, go, -, on date⟩ Positive	I meet a girl. I have a great time.	happy, excited, romantic	go → energy, travel, journey date → lover, engagement
⟨I, have been, -, at hospital⟩ Negative	I was in a car accident. I was released from hospital.	worried, sick, scared	hospital → death, disease, injury
⟨I, save, much money, -⟩ Positive	I work hard at my job. I buy a new car.	happy, satisfied, proud	save → rescue, protect money → rich, reward, earnings
⟨-, separate, child, from family⟩ Negative	Parents go to jail. Mother cries.	sad, unhappy, scared	separate → divorce, abduction child → cute, naughty, noisy family → fellowship, mother
⟨I, have, free weekend, -⟩ Positive	I work all week. I go to the beach to relax.	relaxed, happy, excited	free → fun, gift, independent
⟨I, hear, loud noise, -⟩ Negative	I am walking down the street. I call the police.	scared, alarmed, alert	loud → strong, nightclub, vulgar noise → explosion, bang, trouble

Table 5: Cases that our model makes the correct predictions.

5.3 Case Study

We provide several cases from the Twitter dataset to analyze the influence brought by event contexts, human emotional feedback and sentiment-related concepts. As illustrated in Table 5, the injected knowledge provide interpretable results for the prediction of our model. For events that do not contain sentiment words, such as those listed in Table 5, baseline models tend to classify them as neutral, whereas our model gives the correct predictions. From the cases, it can be observed that the three perspectives of knowledge injection play different roles in sentiment prediction. In most cases, intuitively, context information is of relatively little help in the reasoning process, because the model often does not get direct sentiment-related information from context. However, context information helps the model understand the event, which enriches the original event semantics. Moreover, compared with the other two kinds of information, human emotional feedback brings stronger sentimental signals. Especially when external concepts do not provide obvious sentimental clues, human emotional feedback plays a greater role. Finally, with the help of sentiment-related concepts, the model gains more profound insight into the meaning of words in the event. Since the sentiment of an event is often derived from its predicate and entities, important sentimental clues can be obtained from the extended concepts. Then the implied sentiment can be inferred more easily and more accurately.

6 Conclusion and Future Work

In this paper, we propose a novel Multi-perspective Knowledge-injected Interaction Network (MKIN) for affective event classification. MKIN models various aspects of information by considering contexts, human emotional feedback, and sentiment-related concepts, to fully comprehend the event as well as accurately predict its sentiment. To be more specific, in order to complement the semantic information of the event, we leverage context information and perform context modeling to capture the semantic association between the event and the context. To enhance the sentimental information of the event, we take advantage of human emotional feedback to provide sentimental clues from the perspective of people's emotional state. In addition, external sentiment-related concepts are introduced to enrich the word-level representations. Both emotional state information and concept information fill the reasoning gap between events and sentiments. Experiment results show that knowledge injection from all perspectives improve the model performance, and our model achieves 2.1% performance improvement over the state-of-the-art model on the gold standard dataset.

For future work, to apply the model in a variety of natural language processing applications, we would like to explore event-centered sentiment analysis. Affective event classification can be employed as an additional subtask to improve sentiment analysis.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.

Acknowledgements

We thank the anonymous reviewers for their insightful comments and suggestions. This work was supported by the National Key RD Program of China via grant 2021YFF0901602 and the National Natural Science Foundation of China (NSFC) via grant 62176078.

References

Abien Fred Agarap. 2018. Deep learning using rectified linear units (relu). *CoRR*, abs/1803.08375.

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4762–4779, Florence, Italy, July. Association for Computational Linguistics.

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, June. Association for Computational Linguistics.

Haibo Ding and Ellen Riloff. 2016. Acquiring knowledge of affective events from blogs using label propagation. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Haibo Ding and Ellen Riloff. 2018. Weakly supervised induction of affective events by optimizing semantic consistency. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr.

Amit Goyal, Ellen Riloff, and Hal Daumé III. 2010. Automatically producing plot unit representations for narrative text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 77–86, Cambridge, MA, October. Association for Computational Linguistics.

Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2021. (comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 6384–6392. AAAI Press.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online, July. Association for Computational Linguistics.

Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy. 2014. Major life event extraction from Twitter based on congratulations/condolences speech acts. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1997–2007, Doha, Qatar, October. Association for Computational Linguistics.

Qintong Li, Piji Li, Zhaochun Ren, Pengjie Ren, and Zhumin Chen. 2022b. Knowledge bridging for empathetic dialogue generation. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 10993–11001. AAAI Press.

Jian Liao, Min Wang, Xin Chen, Suge Wang, and Kai Zhang. 2022. Dynamic commonsense knowledge fused method for chinese implicit sentiment analysis. *Inf. Process. Manag.*, 59(3):102934.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Longxiang Liu, Zhuosheng Zhang, Hai Zhao, Xi Zhou, and Xiang Zhou. 2021. Filling the gap of utterance-aware and speaker-aware representation for multi-turn dialogue. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 13406–13414. AAAI Press.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, Melbourne, Australia, July. Association for Computational Linguistics.

Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Takuya Kawada, Stijn De Saeger, Jun'ichi Kazama, and Yiou Wang. 2012. Why question answering using sentiment analysis and word classes. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 368–378, Jeju Island, Korea, July. Association for Computational Linguistics.

Wei Peng, Yue Hu, Luxi Xing, Yuqiang Xie, Yajing Sun, and Yunpeng Li. 2022. Control globally, understand locally: A global-to-local hierarchical graph network for emotional support conversation. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4324–4330. International Joint Conferences on Artificial Intelligence Organization, 7. Main Track.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training.

Sahand Sabour, Chujie Zheng, and Minlie Huang. 2021. Cem: Commonsense-aware empathetic response generation. In *AAAI Conference on Artificial Intelligence*.

Jun Saito, Yugo Murawaki, and Sadao Kurohashi. 2019. Minimally supervised learning of affective events using discourse relations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5758–5765, Hong Kong, China, November. Association for Computational Linguistics.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3027–3035.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607. Springer.

Weiyan Shi and Zhou Yu. 2018. Sentiment adaptive end-to-end dialog systems. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1509–1519, Melbourne, Australia, July. Association for Computational Linguistics.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 4444–4451. AAAI Press.

Elsbeth Turcan, Shuai Wang, Rishita Anubhai, Kasturi Bhattacharjee, Yaser Al-Onaizan, and Smaranda Muresan. 2021. Multi-task learning and adapted knowledge models for emotion-cause extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3975–3989, Online, August. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Hoa Trong Vu, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. 2014. Acquiring a dictionary of emotion-provoking events. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 128–132, Gothenburg, Sweden, April. Association for Computational Linguistics.

Minghao Xu, Daling Wang, Shi Feng, Zhenfei Yang, and Yifei Zhang. 2022a. KC-ISA: An implicit sentiment analysis model combining knowledge enhancement and context features. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6906–6915, Gyeongju, Republic of Korea, October. International Committee on Computational Linguistics.

Ruoxi Xu, Hongyu Lin, Meng Liao, Xianpei Han, Jin Xu, Wei Tan, Yingfei Sun, and Le Sun. 2022b. ECO v1: Towards event-centric opinion mining. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2743–2753, Dublin, Ireland, May. Association for Computational Linguistics.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.

Weixiang Zhao, Yanyan Zhao, and Xin Lu. 2022. Cauain: Causal aware interaction network for emotion recognition in conversations. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4524–4530. International Joint Conferences on Artificial Intelligence Organization, 7. Main Track.

Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 165–176, Hong Kong, China, November. Association for Computational Linguistics.

Deyu Zhou, Jianan Wang, Linhai Zhang, and Yulan He. 2021. Implicit sentiment analysis with event-centered text representation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6884–6893, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Yuan Zhuang, Tianyu Jiang, and Ellen Riloff. 2020. Affective event classification with discourse-enhanced self-training. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5608–5617, Online, November. Association for Computational Linguistics.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 773-785, Harbin, China, August 3 - 5, 2023.