Joint Alignment of Multi-Task Feature and Label Spaces for Emotion Cause Pair Extraction

Shunjie Chen\textsuperscript{1}, Xiaochuan Shi\textsuperscript{1}, Jingye Li\textsuperscript{1}, Shengqiong Wu\textsuperscript{2}, Hao Fei\textsuperscript{2}, Fei Li\textsuperscript{1,\dagger}, Donghong Ji\textsuperscript{1}

\textsuperscript{1}Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China
\textsuperscript{2}School of Computing, National University of Singapore, Singapore

\{shunjiechen, shixiaochuan, theodorelee, lifei_csnlp, dhji\}@whu.edu.cn
swu@u.nus.edu, haofei37@nus.edu.sg

Abstract

Emotion cause pair extraction (ECPE), as one of the derived subtasks of emotion cause analysis (ECA), shares rich inter-related features with emotion extraction (EE) and cause extraction (CE). Therefore EE and CE are frequently utilized as auxiliary tasks for better feature learning, modeled via multi-task learning (MTL) framework by prior works to achieve state-of-the-art (SoTA) ECPE results. However, existing MTL-based methods either fail to simultaneously model the specific features and the interactive feature in between, or suffer from the inconsistency of label prediction. In this work, we consider addressing the above challenges for improving ECPE by performing two alignment mechanisms with a novel A\textsuperscript{2}Net model. We first propose a feature-task alignment to explicitly model the specific emotion&cause-specific features and the shared interactive feature. Besides, an inter-task alignment is implemented, in which the label distance between the ECPE and the combinations of EE&CE are learned to be narrowed for better label consistency. Evaluations of benchmarks show that our methods outperform current best-performing systems on all ECA subtasks. Further analysis proves the importance of our proposed alignment mechanisms for the task.\textsuperscript{1}

1 Introduction

Emotion cause analysis (ECA), detecting potential causes for certain emotion expressions in a document, has been a hot research topic in natural language processing (NLP) community (Lee et al., 2010; Gui et al., 2016; Fan et al., 2019). ECA has derived three associated tasks: EE, CE and ECPE. As illustrated in Figure 1(a), EE and CE detects the emotion and cause clauses respectively, while ECPE identifies both the emotion and cause clauses as well as their semantic relation. By jointly modeling the clauses detection and the relational pairing, ECPE effectively relieves the noise introduction in the pipeline process, and thus receives most research attention recently (Ding et al., 2020a; Wei et al., 2020; Wu et al., 2022).

As there are close correlations among EE, CE and ECPE, existing ECPE works extensively treat EE and CE as auxiliary tasks for additional feature supports, and mostly adopt the multi-task learning framework to explicitly model the inter-dependency in between, thus achieving current SoTA performances (Wei et al., 2020; Ding et al., 2020a; Fan et al., 2021; Bao et al., 2022). From the view of feature encoding, there are two major categories of MTL-based ECPE methods: \textit{shared feature encoding} method and \textit{parallel feature encoding} method. As shown in Figure 1(b), shared

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{(a) illustrates three subtasks of ECA. (b) and (c) depicts the shared and parallel features encoding method, respectively. In (d) and (e) we show our proposed feature-task alignment mechanism and inter-task alignment mechanism, respectively.}
\end{figure}

\textsuperscript{1}Corresponding author

\textsuperscript{1}Our code is available at \url{https://github.com/csj199813/A2Net_ECPE}
methods only learn mixed features via one encoder without distinguishing specific features for individual subtasks (Wei et al., 2020; Yuan et al., 2020). In contrast, parallel methods (Ding et al., 2020a,b; Fan et al., 2021) use two encoders to learn emotion- and cause-specific features separate (cf. Figure 1(c)), where unfortunately, the interaction among these tasks are overlooked. We argue that both the private emotion-specific and cause-specific features and the shared interactive feature are important to the final performance, which should be explicitly modeled in the MTL framework. To this end, we in this work propose a feature-task alignment (FTA) scheme of MTL for ECPE (cf. Figure 1(d)), in which we explicitly split three parts of features, and align them to EE, CE and ECPE respectively.

Meanwhile, aligning the label space in a MTL framework is crucial to overall ECPE performance, because intuitively all modules in the MTL process should reach a consensus. For example as in Figure 1(a), once “c1” is recognized as a cause clause by ECPE module, it should not be further predicted as a non-cause clause by CE module. We notice that such label consistency is not guaranteed in existing MTL-based ECPE methods, which could inevitably hurt the prediction. Therefore we further introduce a inter-task alignment (ITA) mechanism (cf. 1(e)) that learns to pull closer the label distance between the ECPE and the combinations of EE&CE, ensuring label consistency.

We implement the above ideas of feature-task alignment and inter-task alignment by developing a novel neural network, namely A²Net, as shown in Figure 2. First, we employ the BERT (Devlin et al., 2019) as document encoder for producing clauses representations. We then leverage the partition filter network (PFN) (Yan et al., 2021b) to implement the feature-task alignment, generating emotion-specific features, cause-specific features and interaction features, respectively. Afterwards, we apply emotion-specific and interaction features for EE, cause-specific and interaction features for CE, and all features for ECPE. Finally, we reach the goal of inter-task alignment by minimizing the bidirectional KL-divergence between the output distributions of ECPE and EE×CE, thus maintaining the consistency of label spaces among all tasks.

Our A²Net framework is evaluated on the ECA benchmark (Xia and Ding, 2019), where our system achieves new SoTA results on EE, CE and ECPE. Further analyses demonstrate that our method learns better consistency in the predictions of all subtasks than existing baselines. Overall, this work contributes to three major aspects:

- We present an innovative multi-task learning based ECPE framework, where we further propose a feature-task alignment mechanism that can make better use of the shared features from EE and CE sources.
- We also introduce an inter-task alignment mechanism to reduce the inconsistency between the prediction results of ECPE and the EE&CE, significantly enhancing the performance as well as the robustness of the system.
- Our system empirically achieves new SoTA performances of the EE, CE and ECPE tasks on the benchmark.

2 Related Work

In NLP area, the analysis on sentiment and opinion is a long-standing research topic (Liu, 2012; Li et al., 2020; Wu et al., 2021; Fei et al., 2022a), including detecting of the sentiment polarities (Tang et al., 2016; Fei et al., 2022b; Shi et al., 2022) and the emotion categories (Lee et al., 2010; Neviarouskaya and Aono, 2013). One of the recent trend on the emotion detection has been upgraded to the emotion cause analysis (ECA). Centered on the topic of ECA, there are several subordinated tasks according to the extracting elements of emotion and cause, such as emotion cause extraction (ECE) (Lee et al., 2010; Neviarouskaya and Aono, 2013; Gui et al., 2016; Li et al., 2018) and ECPE (Xia and Ding, 2019; Wei et al., 2020; Bao et al., 2022).

Lee et al. (2010) pioneer the ECE task, in which the task is formulated as a word-level cause labeling problem. Following this work, initial research constructs rule-based methods (Neviarouskaya and Aono, 2013; Gao et al., 2015) and machine learning methods (Ghazi et al., 2015; Song and Meng, 2015) on their own corpus. Deep learning based methods greatly facilitate the line of this research (Fei et al., 2021a, 2022e,c, 2021b; Wei et al., 2019a). Recently, Gui et al. (2016) release a public corpus and re-formalize ECE as a clause-level classification problem, where the goal is to detect cause clauses for a given emotion in the text. The framework has received much attention in recent years and the corpus has become a benchmark ECA dataset (Gui et al., 2017; Li et al., 2018; Fan et al., 2019; Ding...
et al., 2019; Hu et al., 2021b; Yan et al., 2021a; Hu et al., 2021a).

However, as Xia and Ding (2019) points out, the ECE task is limited to the task definition, i.e., emotion needs to be manually marked in advance. They thus introduce the ECPE task that simultaneously extract both the emotion and cause clauses as well as determining their relations, which has a better utility in real-world applications (Fei et al., 2022d, 2020). Thereafter, a line of subsequent research efforts are paid to ECPE within the last years (Ding et al., 2020b; Cheng et al., 2020; Fan et al., 2021; Bao et al., 2022).

Recent ECPE methods mostly employ the multi-task learning for simultaneously modeling the EE and CE as auxiliary tasks for making use of the shared features, and thus realize SoTA ECPE performances (Bao et al., 2022). Existing MTL-based ECPE works can largely be divided into two categories: parallel encoding and shared encoding methods. Parallel methods mostly learn the emotion/cause feature representations in mutually independent ways (Cheng et al., 2020; Ding et al., 2020a; Fan et al., 2021; Bao et al., 2022). Ding et al. (2020b); Chen et al. (2022) uses auxiliary task prediction to aid the interactions between emotion and cause features. However, the prediction values are limited to only a two-dimensional vector, leading to insufficient interaction between emotion and cause features.

Shared methods learn the mixed features by only one encoder without distinguishing between features of different tasks (Wei et al., 2020; Yuan et al., 2020; Wei et al., 2019b,c). In this work, we use the PFN (Yan et al., 2021b) to generate emotion-specific features, cause-specific features and interaction features and implement the feature-task alignment. Moreover, we use an inter-task alignment module to reduce the gap between EE, CE and ECPE, maintaining the consistency of the label space.

3 Methodology

Task Formulation Given a document consisting of $N$ clauses $D = \{c_1, c_2, \ldots, c_N\}$, and each $c_i$ denotes a subsection of a sentence separated by a comma. The goals of EE and CE task are extracting emotion clauses $c_i^e \in D$ and cause clauses $c_i^c \in D$ in the document, respectively, while ECPE task identifies the emotion-cause clause pair $(c_i^e, c_i^c)$ that has causal relationship between emotion and cause clauses.

As illustrated in Figure 2, the overall architecture of our $A^2$Net consists of four tiers, including the encoder layer, feature-task alignment layer, prediction layer and inter-task alignment mechanism. First, following the previous work and we use BERT (Devlin et al., 2019) as encoder to yield contextualized clause representations from input documents. Then, to explicitly model task-specific features and task-shared features. We leverage a PFN (Yan et al., 2021b) to capture emotion- and cause-specific features and the interaction between them. Afterward, a prediction layer is used to align three combinations of three kinds of features from PFN with three tasks, and predict the emotion clauses, cause clauses and emotion-cause pair for EE, CE and ECPE, respectively. Finally, considering that there should be a consensus among all tasks. We propose an inter-task alignment mechanism to enhance the consistency between ECPE and EE×CE.

3.1 Encoder Layer

Following (Wei et al., 2020), we also leverage pre-trained BERT language model (Devlin et al., 2019) as the underlying encoder to yield contextualized clause representations. Concretely, we insert a [CLS] token at the beginning of each clause and append a [SEP] token to the end, i.e., $c_i = \{[CLS], w_{i,1}, w_{i,2}, \ldots, w_{i,M}, [SEP]\}$. Then we concatenate them together as the input of BERT to generate contextualized token representations, in which we take the representation of [CLS] token in each clause $c_i$ as its clause representation. After that, we obtain all the clause representations $X = \{x_1, x_2, \ldots, x_N\}$.

3.2 Feature-task Alignment Layer

We adopt partition filter network (PFN) (Yan et al., 2021b) to capture emotion- and cause-specific features and the interaction between them because of its powerful ability to extract task-specific features and interaction features. PFN is similar to the LSTM structure and has two task-related gates: the emotion gate and the cause gate. The gates filter features according to their contribution to each task with emotion and cause gates. In each time step, the encoder divides clause representation into three feature partitions: emotion partition, cause partition, and interaction partition, where interaction partition represents information useful to all tasks.

Specifically, at the $i$ time step, we first generate
two task-related gates:

\[
g_1^c = \text{Cummax}(\text{Linear}([x_i; h_{i-1}])) ,
\]
\[
g_1^e = 1 - \text{Cummax}(\text{Linear}([x_i; h_{i-1}])) ,
\]

where \(\text{Cummax}(\cdot) = \text{Cumsum}(\text{Softmax}(\cdot))\), performs as a binary gate, and \(\text{Linear}(\cdot)\) denotes linear transformation, and \(h_{i-1}\) is the hidden state of \(i\)-1-th clause. Each gate will divide clause representations into two segments: task-related and task-unrelated, according to their usefulness to the specific task. With the joint efforts of the two gates, the clause representation can be divided into three partitions: emotion partition \(p_1^e\), cause partition \(p_1^c\) and interaction partition \(p_1^i\). We use task-related gates \((g_1^c\) and \(g_1^e\)) to calculate forgetting gates:

\[
\begin{align*}
 f_1^i &= g_1^c \circ g_1^e , \\
 f_1^e &= g_1^c - f_1^i , \\
 f_1^c &= g_1^c - f_1^i ,
\end{align*}
\]

where \(\circ\) denotes element-wise multiplication, \(f_1^c\), \(f_1^e\) and \(f_1^i\) are emotion, cause and interaction forgetting gates, respectively. Similarly, we also can obtain input gates \(o_1^i\), \(o_1^e\) and \(o_1^c\) via Equation 1 and 2.

After that, we use forgetting and input gates to control the flow of input and history information:

\[
\begin{align*}
 \tilde{c}_i &= \text{tanh}(\text{Linear}([x_i; h_{i-1}])) , \\
 p_1^i &= f_1^i \circ c_{i-1} + o_1^i \circ \tilde{c}_i , \\
 p_1^e &= f_1^e \circ c_{i-1} + o_1^e \circ \tilde{c}_i , \\
 p_1^c &= f_1^c \circ c_{i-1} + o_1^c \circ \tilde{c}_i ,
\end{align*}
\]

where \(\tilde{c}_i\) denotes the current input information, and \(c_{i-1}\) denotes the history information.

Next, we can obtain three feature representations: emotion feature \(h_1^e\), cause feature \(h_1^c\) and inter-task interaction feature \(h_1^i\) from the partition:

\[
\begin{align*}
 h_1^e &= \tanh(p_1^e) , \\
 h_1^c &= \tanh(p_1^c) , \\
 h_1^i &= \tanh(p_1^i) ,
\end{align*}
\]

We further use the information in all three partitions to construct cell state \(c_i\), and hidden state \(h_i\) for the next time step:

\[
\begin{align*}
 c_i &= \text{Linear}([p_1^e; p_1^c; p_1^i]) , \\
 h_i &= \tanh(c_i) .
\end{align*}
\]

After PFN, we can obtain the emotion-specific features \(h_1^e\), cause-specific features \(h_1^c\) and the interaction features \(h_1^i\). First, we align features with the EE and CE tasks, in which we concatenate the features of emotion and cause with the interaction features separately and gain the emotion representations \(r_1^e = [h_1^e; h_1^c]\) and the cause representations \(r_1^c = [h_1^c; h_1^i]\). Moreover, we consider aligning features with ECPE task, so we add task-specific features and interaction features to aggregate all the information about ECPE task:

\[
\begin{align*}
 h_1^e &= h_1^e + r_1^c , \\
 h_1^c &= h_1^c + r_1^c , \\
 r_{ij} &= [h_1^e; h_1^c']_{ij} ,
\end{align*}
\]

where \(e_{ij}\) denotes the relative position embedding following Wei et al. (2020). \(r_{ij}\) is the final emotion-cause pair representation.
3.3 Prediction Layer

Extracting Emotion/Cause Prediction Layer

We feed \( r^e_i \) and \( r^c_i \) into two feedforward network (FFN) to obtain emotion prediction \( \hat{y}^e_i \) and cause prediction \( \hat{y}^c_i \) for the \( i \)-th clause:

\[
\begin{align*}
\hat{y}^e_i &= \sigma(\text{FFN}(r^e_i)), \\
\hat{y}^c_i &= \sigma(\text{FFN}(r^c_i)),
\end{align*}
\]

where \( \sigma(\cdot) \) means the sigmoid function.

The auxiliary task loss for emotion prediction and cause prediction can be formulated as:

\[
\begin{align*}
&\mathcal{L}_{aux} = -\sum_{i=1}^{N} \left( y^e_i \log(\hat{y}^e_i) + y^c_i \log(\hat{y}^c_i) \right), \tag{8}
\end{align*}
\]

where \( y^e_i \) and \( y^c_i \) are emotion and cause ground truth labels of clause \( c_i \), respectively.

Extracting Emotion Cause Pair

We employ a FFN with a sigmoid function to obtain the emotion-cause joint score \( \hat{y}^{ep}_{ij} \):

\[
\hat{y}^{ep}_{ij} = \sigma(\text{FFN}(r_{ij})). \tag{9}
\]

The loss function of emotion-cause pair extraction can be formalized as:

\[
\mathcal{L}_{pair} = -\sum_{i=1}^{N} \sum_{j=1}^{N} y^e_i y^c_j \log(\hat{y}^{ep}_{ij}), \tag{10}
\]

where \( y^{ep}_{ij} \) is the ground truth label of the clause pair \((c_i, c_j)\).

3.4 Inter-task Alignment Mechanism

As we argued earlier, the predictions of EE and CE could be inconsistent with ECPE, i.e., the emotion in the emotion-cause pair predicted by ECPE could not be detected by EE, which hinders the task for further improvements. To address this issue, we propose an inter-task alignment (ITA) mechanism to constrain the predicted scores between ECPE and auxiliary tasks during the training period. First, we leverage the emotion score \( \hat{y}^e_i \) and the cause score \( \hat{y}^c_j \) to get the pre-pseudo pair score \( \sqrt{\hat{y}^e_i \hat{y}^c_j} \).

Note that there could not exist a causal relationship in pairs matched from predictions of EE and CE. Thus, as shown in Figure 3, we calculate the pseudo emotion-cause pair score \( \hat{y}^{ep}_{ij} \) as follows:

\[
\hat{y}^{ep}_{ij} = \alpha_{ij} \sqrt{\hat{y}^e_i \hat{y}^c_j}, \tag{11}
\]

where \( \alpha_{ij} \) (\( 0 \leq \alpha_{ij} \leq 1 \)) is a soft mask score for the pseudo pair \((\hat{y}^e_i, \hat{y}^c_j)\), which can reduce the score of fake emotion-cause pairs in pre-pseudo scores. \( \alpha_{ij} \) is computed by:

\[
\alpha_{ij} = \frac{\exp(t_{ij})}{\sum_{j}^{N} \exp(t_{ij})}, \tag{12}
\]

where \( v^e_i \) and \( v^c_j \) are obtained from \( r^e_i = [h^e_i; h^*_i] \) and \( r^c_j = [h^c_j; h^*_j] \) with FFNs, respectively. The \( d \) denotes the dimension of \( v^e_i \) and \( v^c_j \).

Then we reduce the gap between the pseudo emotion-cause pair score \( \hat{y}^{ep}_{ij} \) from EE and CE and the true emotion-cause pair score \( \hat{y}^{ep}_{ij} \) from ECPE using Kullback Leibler (KL) Divergence:

\[
\begin{align*}
\mathcal{L}_{KL} &= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\text{KL}(\hat{y}^{ep}_{ij} || \hat{y}^{ep}_{ij}) + \text{KL}(\hat{y}^{ep}_{ij} || \hat{y}^{ep}_{ij})) \\
&= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\hat{y}^{ep}_{ij} \log(\hat{y}^{ep}_{ij} / \hat{y}^{ep}_{ij}) + \hat{y}^{ep}_{ij} \log(\hat{y}^{ep}_{ij} / \hat{y}^{ep}_{ij})). \tag{13}
\end{align*}
\]

Optimization

The final loss of our model is a weighted sum of \( \mathcal{L}_{pair} \), \( \mathcal{L}_{aux} \) and \( \mathcal{L}_{KL} \):

\[
\mathcal{L} = \mathcal{L}_{pair} + \lambda_1 \mathcal{L}_{aux} + \lambda_2 \mathcal{L}_{KL}, \tag{14}
\]

where \( \lambda_1 \) and \( \lambda_2 \) are hyperparameters.
<table>
<thead>
<tr>
<th>Approach</th>
<th>ECPE P</th>
<th>ECPE R</th>
<th>ECPE F1</th>
<th>EE P</th>
<th>EE R</th>
<th>EE F1</th>
<th>CE P</th>
<th>CE R</th>
<th>CE F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTS</td>
<td>72.43</td>
<td>63.66</td>
<td>67.76</td>
<td>81.96</td>
<td>73.29</td>
<td>77.39</td>
<td>74.90</td>
<td>66.02</td>
<td>70.18</td>
</tr>
<tr>
<td>TransECPE</td>
<td>73.74</td>
<td>63.07</td>
<td>67.99</td>
<td>87.16</td>
<td>82.44</td>
<td>87.44</td>
<td>75.62</td>
<td>64.71</td>
<td>69.74</td>
</tr>
<tr>
<td>ECPE-2D</td>
<td>72.92</td>
<td>65.44</td>
<td>68.89</td>
<td>86.27</td>
<td>92.21</td>
<td>90.10</td>
<td>73.36</td>
<td>64.71</td>
<td>71.23</td>
</tr>
<tr>
<td>PairGCN</td>
<td>76.92</td>
<td>67.91</td>
<td>72.02</td>
<td>86.57</td>
<td>79.58</td>
<td>83.75</td>
<td>79.07</td>
<td>69.28</td>
<td>73.75</td>
</tr>
<tr>
<td>RANKCP</td>
<td>71.19</td>
<td>76.30</td>
<td>73.60</td>
<td>91.23</td>
<td>89.99</td>
<td>90.57</td>
<td>74.61</td>
<td>77.88</td>
<td>76.15</td>
</tr>
<tr>
<td>ECPE-MLL</td>
<td>77.00</td>
<td>72.35</td>
<td>74.52</td>
<td>86.08</td>
<td>91.91</td>
<td>88.86</td>
<td>73.82</td>
<td>79.12</td>
<td>76.30</td>
</tr>
<tr>
<td>MGSAG</td>
<td>77.43</td>
<td>73.21</td>
<td>75.21</td>
<td>92.08</td>
<td>82.11</td>
<td>87.17</td>
<td>79.79</td>
<td>74.68</td>
<td>77.12</td>
</tr>
<tr>
<td>A2Net(ours)</td>
<td>75.03</td>
<td>77.80</td>
<td>76.34</td>
<td>90.67</td>
<td>90.98</td>
<td>90.80</td>
<td>77.62</td>
<td>79.20</td>
<td>78.35</td>
</tr>
</tbody>
</table>

Table 1: Comparisons with baselines on Chinese benchmark ECPE corpus. For a fair comparison, they all use BERT as the encoder. The best performance is in **bold** and the second best performance is underlined.

4 Experiments Settings

4.1 Dataset and Metrics

We conducted experiments on the Chinese benchmark dataset provided by Xia and Ding (2019) to evaluate the effectiveness of our proposed model A2Net. For fair comparisons, following previous work we use 10-fold cross-validation as the data split strategy and the precision (P), recall (R), and F1 score (F1) as evaluation metrics. Meanwhile, we also verify the performance of two auxiliary tasks: emotion extraction (EE) and cause extraction (CE), using the same evaluation metrics as ECPE.

4.2 Implementation Details

We apply PyTorch to implement our framework. We leverage pre-trained language model BERT (Devlin et al., 2019) as our embedding layer. We employ one-layer PFN (Yan et al., 2021b) with hidden size of 300. Besides, the hyperparameters $\lambda_1$ and $\lambda_2$ are both set to 0.4. We set the batch size and the learning rate to 4 and 2e-5, respectively. We apply AdamW (Loshchilov and Hutter, 2017) to optimize our model parameters. To prevent overfitting, the dropout rate is set to 0.1.

4.3 Baselines

In order to verify the effectiveness of our proposed model A2Net, we compared it with the following strong methods. For a fair comparison, they all use BERT as the encoder.

- **ANTS** (Yuan et al., 2020): ANTS solves ECPE with a sequence labeling approach and proposes a tagging scheme.
- **TransECPE** (Fan et al., 2020): TransECPE is a transition-based framework that converts ECPE into a parsing-like directed graph construction task.
- **ECPE-2D** (Ding et al., 2020a): ECPE-2D leverages clauses pairs to construct a 2D representation matrix which integrated with auxiliary task predictions for ECPE task.
- **PairGCN** (Chen et al., 2020): This method models the dependency relations among clause pairs with graph convolution networks.
- **RANKCP** (Wei et al., 2020): RANKCP tackles the ECPE task from a ranking perspective and uses graph attention to model the inter-clause relations.
- **ECPE-MLL** (Ding et al., 2020b): ECPE-MLL converts the ECPE task into the emotion-pivot cause extraction and the cause-pivot emotion extraction using the sliding window strategy.
- **MGSAG** (Bao et al., 2022): MGSAG constructs a multi-granularity semantic aware graph to deal with ECPE task, and it is the current SoTA approach.

5 Experimental Results

5.1 Main Results

Table 1 shows the comparison results of our model (A2Net) with the strong baselines on the emotion-cause pair extraction (ECPE) task and two auxiliary tasks: emotion extraction (EE) and cause extraction (CE).
Table 2: Performances (F1) with different feature encoding schemes.

<table>
<thead>
<tr>
<th></th>
<th>ECPE</th>
<th>EE</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A²Net (ours)</td>
<td>76.34</td>
<td>90.80</td>
<td>78.35</td>
</tr>
<tr>
<td>w/ Shared encoding</td>
<td>69.97</td>
<td>84.81</td>
<td>72.66</td>
</tr>
<tr>
<td>w/ Parallel encoding</td>
<td>75.59</td>
<td>89.75</td>
<td>78.03</td>
</tr>
</tbody>
</table>

For the ECPE task, it is clear that our model A²Net achieves 1.13% and 1.82% F1-score improvement over MGSAG (the current best method) and ECPE-MLL, respectively. Further analysis, we can find that the above advantage mainly comes from the improvement of the recall. Compared with MGSAG and ECPE-MLL, our recall is increased by 4.59% and 5.45% respectively, which indicates that consistent prediction on the three tasks allows the A²Net model to detect more emotion-cause pairs under a considerable precision.

For auxiliary tasks, on the EE task, our model achieves 3.63% and 1.94% F1 improvement over MGSAG and ECPE-MLL, and 0.23% F1 improvement over the previous best model RANKCP. On the CE task, our model yields a great improvement of F1 scores by 1.23% in comparison with the top-performing baseline MGSAG, and achieves 2.20% and 2.05% F1 improvement over RANKCP and ECPE-MLL.

We argue that all improvements come mainly from our proposed feature-task alignment module and inter-task alignment module. Both alignment mechanisms are able to collaboratively improve the performances of all tasks, and enhance the robustness of the model. In the following part we performed corresponding experiments to verify our ideas.

5.2 Effect of Feature-task Alignment

The feature-task alignment module is capable of generating efficiency and independent task-specific and interactive features. To verify the effectiveness of our feature-task alignment, we replaced the PFN with two encoding schemes: shared encoding and parallel encoding. The results are shown in Table 2.

In terms of the shared encoding, we encode the emotion features and cause features using a shared BiLSTM, in which emotion and cause features are entangled. For parallel encoding, we utilize two BiLSTMs to capture emotion features and cause features separately, in which interaction informa-

Table 3: Ablation study of inter-task alignment module and auxiliary task (F1). The ECPE→EE×CE means we use the prediction distribution of ECPE to align to EE×CE (i.e., KL(\tilde{y}_{ij}^{p}||\hat{y}_{ij}^{p}) in Eq. 13), and vice versa.

<table>
<thead>
<tr>
<th></th>
<th>ECPE</th>
<th>EE</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A²Net (ours)</td>
<td>76.34</td>
<td>90.80</td>
<td>78.35</td>
</tr>
<tr>
<td>w/o ECPE→EE×CE</td>
<td>75.83</td>
<td>90.65</td>
<td>78.05</td>
</tr>
<tr>
<td>w/o EE×CE→ECPE</td>
<td>75.50</td>
<td>90.54</td>
<td>77.60</td>
</tr>
<tr>
<td>w/o ITA</td>
<td>75.32</td>
<td>90.05</td>
<td>77.37</td>
</tr>
<tr>
<td>w/o EE &amp; CE</td>
<td>74.39</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

5.3 Effect of Inter-task Alignment

In this section, we investigate the effect of the inter-task alignment (ITA) mechanism and auxiliary tasks for A²Net, and the results are plotted in Table 3.

We first analyze the effect of the aligned direction of the inter-task alignment mechanism. When we merely apply unidirectional alignment to regulate the predictions between ECPE and two auxiliary tasks, we can observe slight performance drops on three tasks to some extent. Furthermore, after removing the inter-task alignment mechanism (bidirectional alignment), we find that the overall decreases in F1 score on three tasks happen, and are more than the any unidirectional alignment, which verifies the helpfulness of the alignment in label spaces among tasks, and bi-direction of alignment are more important for our model.

Besides, we also explore the effectiveness of auxiliary tasks, EE and CE. It should be noted that inter-task alignment module does not work after the auxiliary tasks are removed. When the auxiliary tasks are further removed, we can see that the model performances drop significantly, demonstrating that the auxiliary task can effectively contribute to the ECPE task.
On March 14 (c1), a magnitude 4.3 earthquake occurred in Yingquan District, Fuyang City, Anhui (c2). Then (c3), a rumor of a magnitude 6.8 earthquake occurred in Fuyang City at 2:15 am on March 15. (c4), which caused people to panic (c5)... On March 14 (c1), a magnitude 4.3 earthquake occurred in Yingquan District, Fuyang City, Anhui (c2). Then (c3), a rumor of a magnitude 6.8 earthquake occurred in Fuyang City at 2:15 am on March 15. (c4), which caused people to panic (c5)...

Mr. Feng said frankly (c1), Jingjing is naughty on weekdays (c2), and sometimes he is not polite (c3), but when it comes to the reason for the injury (c4), he can’t hide his anger (c5), just because of my son Drank other children’s yogurt (c6). Teacher Xing lost her mind (c7), she was emotionally out of control (c8), then pulled the child out of the door (c9), the child was injured when the door was closed (c10)...

Table 4: Two examples for the case study. The words in orange are the emotion clause, and the words in blue are the cause clause. The green means correct predictions, red means wrong predictions.

<table>
<thead>
<tr>
<th></th>
<th>A^2Net(w/o ITA)</th>
<th>A^2Net</th>
<th>Ground-truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) EE</td>
<td>ECPE: [c7, c6]</td>
<td>ECPE: [c7, c6]</td>
<td>ECPE: [c7, c6]</td>
</tr>
<tr>
<td>b) CE</td>
<td>EE: [c7]</td>
<td>EE: [c7]</td>
<td>EE: [c7]</td>
</tr>
</tbody>
</table>

5.4 Analysis of Prediction Consistency Cross Tasks

In order to verify the effect of our proposed feature-task alignment module and inter-task alignment module on model prediction consistency among tasks, we conduct the experiments with multiple variants of A²Net and the baseline RANKCP, as shown in Figure 4. Specifically, we disassemble the emotion-cause pairs into a set of emotion clauses and a set of cause clauses which are considered as gold labels for EE and CE, respectively. We calculated the consistency rate to evaluate the consistency of EE and CE tasks with ECPE task predictions by (EE & ECPE) / ECPE or (CE & ECPE) / ECPE, where EE, CE and ECPE denote the prediction results of corresponding tasks and & denotes the logic AND.

On the EE task, we can find that when our model A²Net removes the FTA and ITA modules, the consistency rate drops significantly, but when our model only removes FTA or ITA, the consistency rate decreases slightly. Furthermore, all of them outperform RANKCP, which indicates that FTA and ITA have well aligned EE with ECPE.

On the CE task, we can find that a dramatic drop occurs when we remove the FTA and ITA modules. The consistency rate decreases slightly when only ITA is removed, while the consistency rate decreases more when only FTA is removed, indicating that our feature-level alignment is more effective for CE tasks. Moreover, all of them receive better consistency than RANKCP, indicating that both FTA and ITA are able to align CE and

<table>
<thead>
<tr>
<th>Consistency Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>(a) EE</td>
</tr>
<tr>
<td>(b) CE</td>
</tr>
</tbody>
</table>

Figure 4: Consistency of ECPE and EE (a), as well as CE (b).

5.5 Case Analysis

Finally, to better understand the capacity of our proposed model, we empirically perform case study on EE, CE and ECPE tasks. Specifically, we demonstrate some predictions based on three instances randomly selected from testset, as shown in Table 4.

In the first example, our A²Net without ITA correctly predicts the emotion cause pair (c7, c6) on the ECPE task and incorrectly on the EE and CE tasks. In contrast, our A²Net model correctly predicts all ECPE, EE, and CE tasks after going through the inter-task alignment module. In the second example, A²Net without ITA correctly predicts the emotion cause pair (c5, c4) on the ECPE task and incorrectly on the CE tasks. However, A²Net model correctly predicts all EE, CE and
Table 5: Parameter number and inference speed comparisons on ECPE. All models are tested with batch size 4.

ECPE tasks. In the third example, $A^2$Net without ITA predicts correctly in the EE task and incorrectly detects the cause clause $c_0$ in the CE task. Meanwhile, the emotion-cause pair $(e_0, c_0)$ is predicted incorrectly. Nonetheless, after the inter-task alignment module, all predictions were correct in both the ECPE and CE tasks.

This shows that after aligning between tasks, the model can identify emotion-corresponding causes, which is like the role of our proposed soft mask score. We find that these cases are common in our dataset, which ultimately leads directly to an improvement in our model performance. This also demonstrates that our inter-task alignment module can normalize the inter-task training to make the model performs better and more stable.

5.6 Model Efficiency Analysis

In order to test the efficiency of our model, we conduct the experiments with multiple variants of $A^2$Net and the baseline RANKCP, and the results are plotted in Table 5.

In terms of parameter quantity, our $A^2$Net model is even less than RANKCP, and it can be found that since the ITA module does not introduce additional parameters, w/o ITA does not change the parameter quantity. In terms of inference speed, our model has the same inference speed as RANKCP, which shows that the efficiency of our model does not decrease due to the addition of the alignment mechanism.

6 Conclusion

Existing best-performing ECPE works extensively leverage EE and CE as auxiliary tasks for better feature learning via multi-task learning (MTL). In this paper, we further enhance the existing best-performing MTL-based ECPE by proposing feature-task alignment and inter-task alignment mechanisms. At the feature space, the feature-task alignment mechanism aligns the task-specific features and the shared interactive feature with corresponding tasks. At the label space, the inter-task alignment mechanism reduces the inconsistency among the predicted labels of EE, CE and ECPE. Experimental results on the benchmark ECPE data demonstrate the effectiveness of our methods. Further analysis shows that our system achieves better consistency than existing baselines, which explains the improvements of our model. The idea to align the feature space and label space in MTL framework is promising. In the future work, we consider further constructing intra-clause relations, inter-clause relations, and relations among pairs of clauses for ECPE.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 62176187), the National Key Research and Development Program of China (No. 2017YFC1200500), the Research Foundation of Ministry of Education of China (No. 18JZD015), the Youth Fund for Humanities and Social Science Research of Ministry of Education of China (No. 22YJCZH064), the General Project of National Science Foundation of Hubei Province (No.2021CFB385). This work is also the research result of the independent scientific research project of Wuhan University, supported by the Fundamental Research Funds for the Central Universities.

References


