An End-to-end Mutually Interactive Emotion-Cause Pair Extractor

09118120 Haoqing Xu[1], 09118102 Yan Zhang[1], 09118104 Xiao Tan[1]

[1] *Southeast University, Nanjing, Jiangsu, China*

June 16, 2021

Emotion-cause pair extraction (ECPE), the task aimed at finding the clause-pair of emotion and cause expressions in a piece of document, has gained great population and application in the recent years. Without the given annotated emotion word, ECPE can automatically find the corresponding clause of the emotions in the text. However, previous ECPE related models either fail to develop an end-to-end model, which ignores the link between emotion & cause extraction and emotion-cause pairing, or face huge computational cost when matching the emotion-cause pair. In this paper, we propose an end-to-end mutually interactive emotion-cause pair extractor (Emiece) that uses two soft-shared LSTMs in an end-to-end model to measure the weighted probability of being potential emotion-cause pairs. We conduct experiments on an existing Emotion-Cause Extraction (ECE) corpus introduced in the NTCIR-13 Workshop (Gao et al., 2017). Our model achieves great improvement compared with the original two-step ECPE model and has advantage over other end-to-end models. The experiment results show that Emiece is comparable to the state-of-the-art ECPE models.

1 Introduction

Recently, emotion-cause extraction (ECE) has gained great popularity in text analysis (Lee, Chen, and Huang, 2010; Russo et al., 2011). ECE aims at extracting potential causes that lead to emotion expressions in the text. Instead of using word-level labeled sequence, ECE concentrates on the clause-level sequence, thus fully exploiting the linked relationship between different sentences (Gui et al., 2018). This kind of clause-level observation improves the reliability of the ECE analysis greatly. In this respect, (Gui et al., 2016) first released a corresponding corpus and has been widely used in the following study. However, the ECE task has two limitations: 1) The emotions must be given in the test set, the annotation hugely wastes massive resources and limits the application of this task at the same time. 2) ECE ignores the mutual relationship between the emotion and the cause clause in the text.

To solve the existing problems, (Xia and Ding, 2019) proposed a new task, emotion-cause pair extraction (ECPE), and designed a two-step LSTM model to solve the task. Without the given annotated emotion word, it can automatically find the corresponding clause of the emotions in the text. Figure 1 shows an example. ECPE directly extracts the emotion clause “The old man was very happy" and its corresponding cause clauses “a policeman visited the old man with the lost money", “and told him that the thief was caught". ECPE only needs raw documents , so it saves much time in manual annotating. In addition, ECPC can generate all the emotion-cause pairs in the text so that it offers a more comprehensive understanding of the context. Therefore, it can be widely applied in real-time sentiment-cause analysis in twitter, facebook, comments on goods, films and so on.

Despite ECPE contributes much in comprehensive emotion-cause analysis, the separated two steps still make it impossible to train an end-to-end model, so the cumulative error in the first step will affect the result in the second step. Although the previous models (Singh et al., 2021; Xia and Ding, 2019) consider the mutual interaction between emotion and cause, the process is simply finished by transferring information from one decoder to another uniaxially, thus harming the mutual information transfer. Moreover, (Xia and Ding, 2019) applies grid searching to finding the best emotion-cause pair, so it may lead to huge computa-

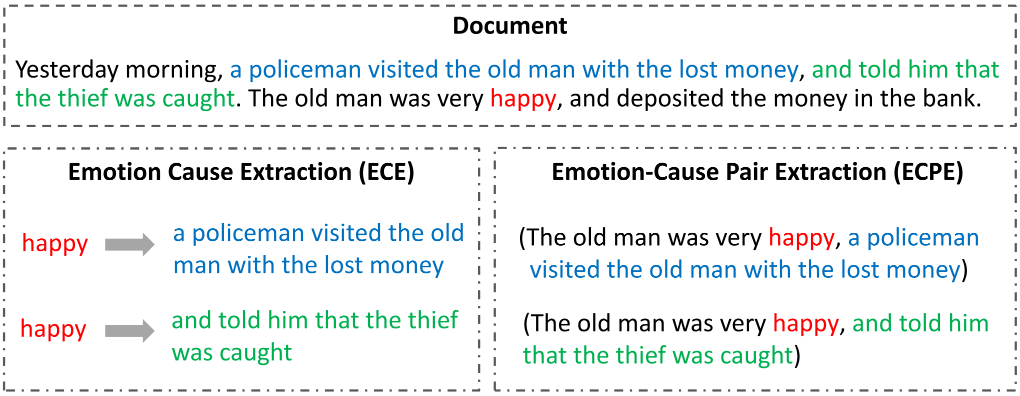

Figure 1: *An example showing the difference between the ECE task and the ECPE task.*

tional cost when facing a large amount of text information. Therefore, how to efficiently extract emotion-cause pairs from all the potential clause pairs is an important problem in ECPC since the pair number grows much quickly when dealing with long documents.

In order to address the above challenges, we propose an end-to-end model, `Emiece`, that predicts the emotion-cause pair from the raw document. We apply soft-sharing between the emotion encoder and the cause encoder to mutually transfer the information extracted from clause representations. The each clause is passed through the emotion and cause detectors to generate emotion and cause clause probabilities that indicate the importance of the clause in terms of emotion and cause. Thus, we use them in the higher layer of the model to predict the emotion-cause pairs.

The main contributions of our work can be summarized as follows:

- **Novel end-to-end ECPE model.** We propose a novel end-to-end method, `Emiece` that uses two LSTMs to automatically transfer information between emotion and cause encoder. Since the end-to-end model considers single emotion and cause extraction along with emotion-cause pairing at the same time, it greatly avoids the cumulative errors in separated steps and improves the performance a lot.
- **Mutual Information Transfer in Emotion and Cause Extraction.** Soft-sharing is applied between emotion and cause encoders. We add the soft-sharing loss to the total loss function in a multi-task learning style in order to involve mutual interaction between the two auxiliary tasks. Therefore, the two encoders can learn from each other rather than unidirectional learning in previous methods.
- **Efficient pair extractor with weighted representation.** We use the weighted representation of emotion and cause to filter the clauses which tend to be meaningless. Therefore, only the useful emotion-weighted and cause-weighted clause representations can be reserved to improve the efficiency of emotion-cause paring.
- **Experiments on large English-language corpus.** We conduct our experiments on a ECPE task suitable corpus based on the benchmark English-language corpus used in the ECE task of the NTCIR-13 workshop (Gao et al., 2017). We compare the performance of both ECE task and ECPE task and find that our model outperforms previous ECE related models.

2 Related Work

The emotion-cause extraction (ECE) task was first proposed in (Lee, Chen, and Huang, 2010). As a word-level task, the extraction is fulfilled with traditional machine learning and rule-based approaches (Chen et al., 2010; Gao, Xu, and Wang, 2015; Li and Xu, 2014). For example, in (Li and Xu, 2014), the authors proposed a fine-grained rule-based method for the task and conducted experiments on the Chinese microblog posts corpus labeled by human annotators. Despite the overall performance of word-level task is not promising enough, it provides a new way to look at the emotion classification task.

Another kind of emotion-cause extraction task is based on clause that solves the problem of word-level labeling in previous work (Serban et al., 2017; Bahdanau, Cho, and Bengio, 2014; Sutskever, Vinyals, and Le, 2014; Schuster and Paliwal, 1997). In (Serban et al., 2017), the authors employed a multi-kernel learning method for the clause-level task on a Chinese emotion cause corpus. Moreover, as the development of deep learning, multiple recurrent neural network (RNN) related models have been proposed to solve clause-level tasks due to its excellent performance in analyzing the relationship between different sequences. Long short-term memory (LSTM), an advanced version of RNN, achieves better performance in related tasks thanks to its forgetting mechanism (Schuster and Paliwal, 1997). Although clause-level methods relax word-level annotations into clause-level annotations and achieve higher performance due to the development of neural networks, it is still restricted by the manual annotations. In addition, it ignores the mutual relationship and interaction between emotion and cause.

To overcome the mentioned drawbacks of ECE, (Xia and Ding, 2019) proposed emotion-cause pair extraction (ECPE) task and they conducted a 2-step hierarchical structure network for the task. This model separates the emotion & cause extraction and the pairing into two steps, therefore, the mistakes made in the first step will affect the results of the second step. To solve these limitations in (Xia and Ding, 2019), several ECPE models have been proposed (Ding, Xia, and Yu, 2020a; Ding, Xia, and Yu, 2020b; Singh et al., 2021). The comparison between these ECPE methods and our proposed method is mentioned in section 4.

3 Methodology

3.1 Task Formalization

Formally, a document consists of text that is segmented into an ordered set of clauses $D = [c_1, c_2, ..., c_d]$ and the ECPE task aims to extract a set of emotion-cause pairs $P = \{..., (c_i, c_j), ...\}$ $(c_i, c_j \in D)$, where c_i is an emotion clause and c_j is the corresponding cause clause.

3.2 Architecture

The whole model contains three layers as illustrated in Figure 2: word-level encoder layer, clause-level encoder layer and pairing layer. We take the vector representation $\boldsymbol{v}_{ij}$ of the j-th word in the i-th clause as input. For each clause, the word vector sequence $\{\boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, \ldots, \boldsymbol{v}_{i,m}\}$ is passed through a word-level encoder, implemented via a Bi-LSTM with attention (Bahdanau, Cho, and Bengio, 2014). The word-level encoder outputs a clause representation $\boldsymbol{s}_i$ for each clause. The higher level contains two clause-level encoders implemented via the stacked Bi-LSTM (Graves, Mohamed, and Hinton, 2013) for emotion clause detection and cause clause detection, respectively. The two encoders take the clause representation sequence $\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_d\}$ as input and generate the emotion and cause representation of the clauses $\boldsymbol{r}_i^e$, $\boldsymbol{r}_i^c$. In order to mutually transfer the information obtained by the encoders, we use a soft-sharing strategy between the two encoders. The representations are then fed into detectors (logistic regression layers) to get the probability distribution $\boldsymbol{a}_i^e$, $\boldsymbol{a}_i^c$ of the clause being an emotion clause and a cause clause, respectively, formed as

$$\begin{aligned}\boldsymbol{a}_i^e &= \mathrm{softmax}(\boldsymbol{W}^e \boldsymbol{r}_i^e + \boldsymbol{b}^e) \\ \boldsymbol{a}_i^c &= \mathrm{softmax}(\boldsymbol{W}^c \boldsymbol{r}_i^c + \boldsymbol{b}^c)\end{aligned} \tag{1}$$

where $\boldsymbol{W}^e$, $\boldsymbol{W}^c$, $\boldsymbol{b}^e$, $\boldsymbol{b}^c$ are the parameters for emotion and cause detection layers. We denote a_i^e, a_i^c for the elements in $\boldsymbol{a}_i^e$, $\boldsymbol{a}_i^c$ that stand for the probabilities of true emotion and cause clause, respectively.

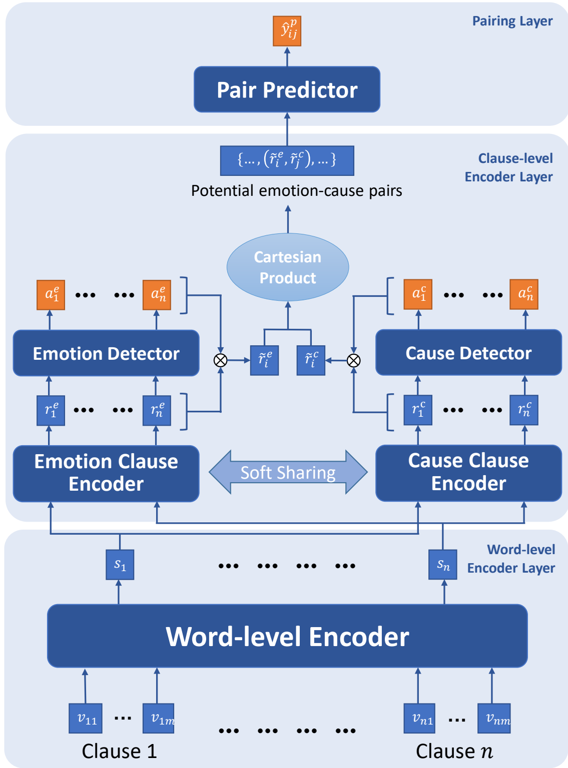

Figure 2: *An illustration of our proposed end-to-end mutually interactive model.*

These probabilities can be viewed as the attentions of emotion and cause encoders. Thus, we multiply the clause representation by the probabilities to obtain the emotion-weighted and cause-weighted clause representation $\widetilde{\boldsymbol{r}}_i^e$, $\widetilde{\boldsymbol{r}}_i^c$ as

$$\begin{aligned}\widetilde{\boldsymbol{r}}_i^e &= a_i^e \boldsymbol{r}_i^e \\ \widetilde{\boldsymbol{r}}_i^c &= a_i^c \boldsymbol{r}_i^c\end{aligned} \tag{2}$$

Once we collect all the weighted representations of clauses into two sets $\mathcal{E} = \{\widetilde{\boldsymbol{r}}_1^e, \widetilde{\boldsymbol{r}}_2^e, \cdots, \widetilde{\boldsymbol{r}}_d^e\}$ and $\mathcal{C} = \{\widetilde{\boldsymbol{r}}_1^c, \widetilde{\boldsymbol{r}}_2^c, \cdots, \widetilde{\boldsymbol{r}}_d^c\}$, we apply the Cartesian product on the two sets to generate all the potential emotion-cause pairs $(\widetilde{\boldsymbol{r}}_i^e, \widetilde{\boldsymbol{r}}_i^c)$. We use $\boldsymbol{r}_{ij}^p = \widetilde{\boldsymbol{r}}_i^e \oplus \widetilde{\boldsymbol{r}}_i^c \oplus \boldsymbol{d}_{ij}$ as the representation of a pair, where $\oplus$ denotes the concatenation operator and $\boldsymbol{d}_{ij}$ is the positional embedding vector representing the relative positioning between clause i and j (Shaw, Uszkoreit, and Vaswani, 2018). The pairs are fed into the pairing layer one at a time to obtain the predicted label. The pairing layer is a fully-connected layer as

$$\begin{aligned}\boldsymbol{h}_{ij} &= \mathrm{ReLU}(\boldsymbol{W}^h \boldsymbol{r}_{ij}^p + \boldsymbol{b}^h) \\ \boldsymbol{y}_{ij}^p &= \mathrm{softmax}(\boldsymbol{W}^y \boldsymbol{h}_{ij} + \boldsymbol{b}^y)\end{aligned} \tag{3}$$

Here, $\boldsymbol{y}_{ij}^p$ gives the Bernoulli distribution probabilities of (c_i, c_j) to be a emotion-cause pair. In total, there are three tasks: one primary task for predicting pairs and two auxiliary tasks for predicting emotion and cause clauses. The outputs are $\boldsymbol{y}_{ij}^p$, $\boldsymbol{a}_i^e$, $\boldsymbol{a}_i^c$, respectively.

3.3 Training

In order to train the end-to-end model, we set the loss function to be

$$L_{\text{total}} = \lambda_c L_c + \lambda_e L_e + \lambda_p L_p + \lambda_{sf} L_{sf} \tag{4}$$

where L_c, L_e and L_p are the cross-entropy errors of emotion clause detection, cause clause detection and pair extraction, respectively, L_{sf} is the soft-sharing loss.

Following (Guo, Pasunuru, and Bansal, 2018), we define

$$L_{sf} = \sum_{d \in \mathcal{D}} \|\phi_d^e - \phi_d^c\|^2 \tag{5}$$

where $\mathcal{D}$ is the set of sharing parameter indices, ϕ^e and ϕ^c are the emotion and cause encoder parameters, respectively. The works in the recent years find (Howard and Ruder, 2018; Yosinski et al., 2014) that features extracted in the shallow layers of a deep neural network contains more general features of different tasks. The higher the layer, the more task-specific the features are extracted. Following the work of (Guo, Pasunuru, and Bansal, 2018), we apply the similar soft-sharing strategy on the first-layer Bi-LSTM in the emotion and cause encoders.

To avoid the unbalance of positive pairs and negative pairs in the pair extraction, the loss L_p is rewised as

$$L_p = L_p^+ + \lambda_- L_p^- \tag{6}$$

where L_p^+, L_p^- denotes the term of positive and negative ground truths in the cross-entropy loss function. λ_- is relatively small since the negative pairs is much fewer than positive ones.

4 Experiments

4.1 Dataset

The dataset was constructed by (Singh et al., 2021) from an existing Emotion-Cause Extraction (ECE) corpus. The corpus was introduced in the NTCIR-13 Workshop (Gao et al., 2017) for ECE challenge.

There are 2,843 documents taken from several English novels and each document is annotated with: 1) emotion-cause pairs (the set of emotion clauses and their corresponding cause clauses); 2) emotion category of each clause; 3) keywords in the emotion clauses. In our experiments, we do not use the emotion category or keywords and only exploits the emotion-cause pairs in training process. None of the annotations are used when testing the model. The whole dataset is split by 80%-10%-10% for training, validating, and testing. 10 such random splits are generated and average results are reported.

4.2 Baselines and Settings

We include three baseline methods in our comparison with `Emiece`: the original two-step model `ECPE` (Xia and Ding, 2019), a state-of-the-art ECPE model `ECPE-2D` (Ding, Xia, and Yu, 2020a), and another end-to-end model $\texttt{E2E-PExt}_\texttt{E}$ (Singh et al., 2021). Our model is trained for 15 epochs using Adam optimizer (Kingma and Ba, 2014). We set the learning rate $\alpha = 0.005$, batch size $N = 32$. The model parameters are initialized randomly following uniform distribution $U(-0.1, 0.1)$. We use GloVe word embedding (Pennington, Socher, and Manning, 2014) of 200 dimension. We set the dropout rate to 0.8 for word embeddings and ℓ_2 decay of 10^{-5} on softmax parameters. The loss weights are set as $\lambda_e : \lambda_c : \lambda_p : \lambda_s f = 1 : 1 : 2.5 : 0.25$ and the negative pair weight $\lambda_- = 0.4$. To obtain better positional embeddings which encode the relative positioning between clauses, we trained randomly initialized embeddings after setting the clipping distance (Shaw, Uszkoreit, and Vaswani, 2018) to 10.

4.3 Metrics

Following (Xia and Ding, 2019), we use precision, recall and F1-score as metrics to evaluate the performance of each model. These metrics are defined as:

$$\text{Precision} = \frac{\#correct\_pairs}{\#proposed\_pairs}$$

$$\text{Recall} = \frac{\#correct\_pairs}{\#annotated\_pairs}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where $\#correct\_pairs$ is the number of correctly predicted emotion-cause pairs, $\#proposed\_pairs$ is the number of all the predicted pairs, $\#annotated\_pairs$ is the number of true pairs in the documents. The definition for the three metrics in two auxiliary tasks is similar to that above.

4.4 Results

The results are summarized in Table 1. Our method achieves a significant improvement of about 7% in pair extraction task compared with the original `ECPE` model. This suggests that an end-to-end framework does have advantage in the primary task. Additionally, our model outperforms $\texttt{E2E-PExt}_\texttt{E}$, an end-to-end model with unidirectional information transfer, in cause extraction and pair extraction, with a neglectable drop in emotion extraction. It means that our soft-sharing strategy benefits the model on the two tasks more efficiently than one-way information transferring. Comparing our model with the state-of-the-art model `ECPE-2D`, there is only a tiny gap in pair extraction task and comparable performance in auxiliary tasks.

	Emotion Extraction			Cause Extraction			Pair Extraction		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
ECPE	67.41	**71.60**	69.40	60.39	47.34	53.01	46.94	41.02	43.67
ECPE-2D	74.35	69.68	**71.89**	64.91	**53.53**	**58.55**	**60.49**	43.84	**50.73**
E2E-PExt$_E$	71.63	67.49	69.43	66.36	43.75	52.26	51.34	**49.29**	50.17
Emiece (Ours)	**75.16**	64.85	69.38	**69.22**	44.26	53.80	54.38	47.73	50.70

Table 1: *Experiment results of our method and baselines. The bold numbers stand for the best performance of the four and the underlined numbers stand for the second best.*

5 Conclusion

In this paper, we propose an end-to-end model that mutually transfers information via soft sharing between emotion and cause extraction tasks. We demonstrate that such interaction and an end-to-end framework benefit the model performance in the primary pair extraction task. Evaluation on the dataset shows that our model has comparable performance with the state-of-the-art models and outperforms the original two-step model. In the future, we plan to explore other approaches to efficient use of the data and richer representations for semantic extraction tasks.

References

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2014). "Neural machine translation by jointly learning to align and translate". In: *arXiv preprint arXiv:1409.0473*.

Chen, Ying et al. (2010). "Emotion cause detection with linguistic constructions". In: *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 179–187.

Ding, Zixiang, Rui Xia, and Jianfei Yu (2020a). "ECPE-2D: Emotion-Cause Pair Extraction based on Joint Two-Dimensional Representation, Interaction and Prediction". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3161–3170.

– (2020b). "End-to-End Emotion-Cause Pair Extraction Based on Sliding Window Multi-Label Learning". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3574–3583.

Gao, Kai, Hua Xu, and Jiushuo Wang (2015). "Emotion cause detection for chinese micro-blogs based on ecocc model". In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pp. 3–14.

Gao, Qinghong et al. (2017). "Overview of NTCIR-13 ECA task". In: *Proceedings of the NTCIR-13 Conference*.

Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton (2013). "Speech recognition with deep recurrent neural networks". In: *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee, pp. 6645–6649.

Gui, Lin et al. (2016). "Emotion cause extraction, a challenging task with corpus construction". In: *Chinese National Conference on Social Media Processing*. Springer, pp. 98–109.

Gui, Lin et al. (2018). "Event-driven emotion cause extraction with corpus construction". In: *Social Media Content Analysis: Natural Language Processing and Beyond*. World Scientific, pp. 145–160.

Guo, Han, Ramakanth Pasunuru, and Mohit Bansal (2018). "Soft layer-specific multi-task summarization with entailment and question generation". In: *arXiv preprint arXiv:1805.11004*.

Howard, Jeremy and Sebastian Ruder (2018). "Universal language model fine-tuning for text classification". In: *arXiv preprint arXiv:1801.06146*.

Kingma, Diederik P and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Lee, Sophia Yat Mei, Ying Chen, and Chu-Ren Huang (June 2010). "A Text-driven Rule-based System for Emotion Cause Detection". In: *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Los Angeles, CA: Association for Computational Linguistics, pp. 45–53. URL: https://www.aclweb.org/anthology/W10-0206.

Li, Weiyuan and Hua Xu (2014). "Text-based emotion classification using emotion cause extraction". In: *Expert Systems with Applications* 41.4, pp. 1742–1749.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning (2014). "Glove: Global vectors for word representation". In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Russo, Irene et al. (June 2011). "EMOCause: An Easy-adaptable Approach to Extract Emotion Cause Contexts". In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*. Portland, Oregon: Association for Computational Linguistics, pp. 153–160. URL: https://www.aclweb.org/anthology/W11-1720.

Schuster, Mike and Kuldip K Paliwal (1997). "Bidirectional recurrent neural networks". In: *IEEE transactions on Signal Processing* 45.11, pp. 2673–2681.

Serban, Iulian et al. (2017). "A hierarchical latent variable encoder-decoder model for generating dialogues". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. 1.

Shaw, Peter, Jakob Uszkoreit, and Ashish Vaswani (2018). "Self-attention with relative position representations". In: *arXiv preprint arXiv:1803.02155*.

Singh, Aaditya et al. (2021). "An End-to-End Network for Emotion-Cause Pair Extraction". In: *arXiv preprint arXiv:2103.01544*.

Sutskever, Ilya, Oriol Vinyals, and Quoc V Le (2014). "Sequence to sequence learning with neural networks". In: *arXiv preprint arXiv:1409.3215*.

Xia, Rui and Zixiang Ding (2019). "Emotion-cause pair extraction: A new task to emotion analysis in texts". In: *arXiv preprint arXiv:1906.01267*.

Yosinski, Jason et al. (2014). "How transferable are features in deep neural networks?" In: *arXiv preprint arXiv:1411.1792*.