Combining Event-level and Cross-event Semantic Information for Event-Oriented Relation Classification by SCNN

Siyuan Ding, Yu Hong, Shanshan Zhu, Jianmin Yao, and Qiaoming Zhu

Provincial Key Laboratory of Computer Information Processing Technology,
Soochow University, Suzhou, China
{dsy.ever,zhushanshan063,tianxianer}@gmail.com,
jmyao@szkj.gov.cn,qmzhu@suda.edu.cn

Abstract. Previous researches on event relation classification primarily rely on lexical and syntactic features. In this paper, we use a Shallow Convolutional Neural Network (SCNN) to extract event-level and cross-event semantic features for event relation classification. On the one hand, the shallow structure alleviates the over-fitting problem caused by the lack of diverse relation samples. On the other hand, the utilization and combination of event-level and cross-event semantic information help improve relation classification. The experimental results show that our approach outperforms the state of the art.

Keywords: Event Relation Classification; Semantic Information; Frame Embedding; SCNN

1 Introduction

The task of Event Relation Detection (abbr., ERD) is defined to determine semantic relation between event mentions, and aggregate distributed events in text to form event relation network. ERD is comprised of two separate tasks, Event Relation Identification (ERI) and Event Relation Classification (ERC) [14]. ERI intends to determine whether two events are relevant. ERC determines what types of relations occurred between the events. In this paper, we take our research focus on ERC.

The input of an ERC system is a pair of event mentions. An event mention is a sentence or a clause that depicts a natural event, consisted of at least the trigger of the event and the closely related participants. See the following event mentions, for example, which are respectively by the trigger words *attacked* and *wounded and died*:

Event1: *Terrorists attacked Bataclan Theatre,*
Event2: *many people wounded and died.*

The output is a tag of relation type, such as that between the above mentions, `Causality`, which inherits the main relation type `Contingency`. As shown in Table 1, there are four top-relations in the first level and ten sub-relations in the second level. In this paper, four top-relation types wiil be considered for the evaluation of ERC systems, including `Contingency`, `Expansion`, `Comparison` and `Temporality`.

Table 1. The architecture of event relations

	Top-relation	**Sub-relation**
Relations	Comparison	Concession
		Contrast
	Contingency	Cause
		Condition
	Expansion	Instantiation
		List
		Progression
		Restatement
	Temporal	Asynchronous
		Synchronous

This paper shows a pilot study on CNN based ERC. Our goal is to introduce semantic-level relation analysis into the perception of logical relation among real historical events. In particular, we embed cross-event semantic features, along with inner ones of a single relation sample. By combination of the features, we can deal with the relation classification for the non-adjacent event mentions and even cross-document and cross-topic samples, such as the `Comparison` relation between the events "*tsunami alarm in Hawaii*" and "*Many planes turn back to San Francisco*".

2 Related Work

Pattern-matching method is one of the conventional approaches on ERC. Chklovski and Pantel [4] extracts pairwise events on the basis of manual designed lexical-syntactic pattern. Pantel and Pennacchiotti [11] propose a method based on Espresso Algorithm to construct patterns automatically, which somewhat improves the recall of pattern-matching method.

The most recent research takes event elements as the clues for relation inference. They are mainly inherit Harris distribution assumption that words in the same context usually hold the same or similar meaning [9]. Lin and Pantel [10] propose an unsupervised method relying on Harris assumption and dependency tree. The algorithm identifies grammatical relationships between words and constructs dependency trees formed of the relationships between words.

Ding et al. [6] present a semi-supervised approach based on Tri-Training. Though these methods have been proven successful, manual features are still weak in capturing semantic aspects of events.

Recently, Zhang et al. [15] succeed in using SCNN to implicit discourse relation recognition, which considerably promotes the development of relation detection tasks. However, we find that the method isolates the inner clues between

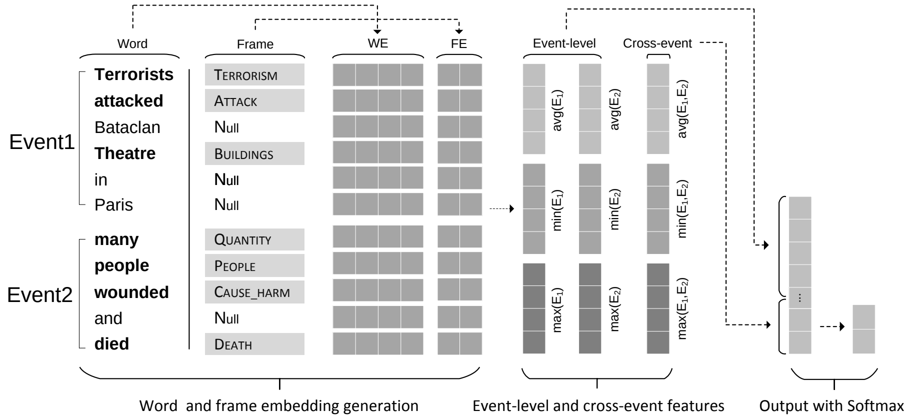

Fig. 1. The architecture for event relation classification via SCNN by combining event-level and cross-event features. WE=word embedding and FE=frame embedding.

two arguments and only takes each argument into consideration, which inevitably omits the mutual influence between two arguments.

3 Methodology

Figure 1 describes the architecture of our model, which primarily involves the following 4 components: (1) word embedding learning, which reveals the embedding vector of words in event; (2) frame embedding generation, where we extract frames from event to generate the frame embedding vector; (3) event-level and cross-event features extraction with SCNN, which exploits the inner semantics in and cross event; and (4) relation classification, which concatenates the generated features and outputs the candidate relation with highest confidence score.

3.1 Word Embedding Learning

We choose the state-of-the-art model Skip-gram to pre-train the word embedding [2]. In our framework, $w_i \in \mathrm{R}^d$ corresponds to a d-dimensional vector representation of i-th word in each event,

$$w_i = (x_i^1, x_i^2, ..., x_i^d) \tag{1}$$

where x_i^j denotes the real-value of j-th dimension in i-th word's embedding, $1 \leq j \leq d$. Figure 1 assumes that each word has size $d = 4$.

3.2 Frame Embedding Generation

The frame semantics information in FrameNet [7] is proved effective to many natural language processing problems [1, 3, 12, 14].

Frame semantics conceptualizes those events with same or similar semantic properties. If two event-event pairs share the similar scenario (which is comprised of a series of frames), they always hold the same relation [14]. Therefore, we use frame semantics as another important information to help detect event relations.

As shown in Figure 1, we use SEMAFOR[1] to identify the frames of each event. Then, the frames will be mapped to a randomly initialized vector of dimension l, let $f_i \in \mathrm{R}^l$ correspond to a l-dimensional vector of i-th frame:

$$f_i = (x_i^1, x_i^2, ..., x_i^l) \tag{2}$$

3.3 Extracting Event-level and Cross-event Features with SCNN

We combine word embedding w_i and frame embedding f_i to generate e_i,

$$e_i = w_i \oplus f_i \tag{3}$$

where $\oplus$ is a concatenation operator, $e_i \in \mathrm{R}^{d+l}$. An event with n words is extracted as the following matrix,

$$E = ({e_1}^T, {e_2}^T, ..., {e_n}^T)^T \tag{4}$$

where $E \in \mathrm{R}^{n \times (d+l)}$. [15] follows previous works [5, 13] and explores three convolutional operations in detecting discourse relations. We adopt this SCNN method and for each column c in E take the following three convolution operations to capture event-level features,

• **event-level features**:

$$max(E^c) = \max\{e_1^c,\ e_2^c,\ ...,\ e_n^c\} \tag{5}$$

$$min(E^c) = \min\{e_1^c,\ e_2^c,\ ...,\ e_n^c\} \tag{6}$$

$$avg(E^c) = \frac{1}{n}\sum_i^n e_i^c \tag{7}$$

where $1 \leq c \leq d$, e_i^c refers to the value of c-th dimension in e_i of event E. For each column c, we further obtain the cross-event features by exploring *max*, *min* and *avg* operations on matrix $\{E_1, E_2\}$,

• **cross-event features**:

$$max(E_1^c,\ E_2^c) = \max\{max(E_1^c),\ max(E_2^c)\} \tag{8}$$

$$min(E_1^c,\ E_2^c) = \min\{min(E_1^c),\ min(E_2^c)\} \tag{9}$$

$$avg(E_1^c,\ E_2^c) = \frac{1}{2}\{avg(E_1^c),\ avg(E_2^c)\} \tag{10}$$

[1] http://www.ark.cs.cmu.edu/SEMAFOR

In the next step, an concatenation operation is performed on E_1, E_2 and $\{E_1, E_2\}$ to generate $a(E_1)$, $a(E_2)$ and $a(E_1, E_2)$,

$$a(E) = max(E) \oplus min(E) \oplus avg(E) \tag{11}$$

where $a \in \mathrm{R}^{3\times(d+l)}$. After getting event-level and cross-event convolution features, we concatenate $a(E_1)$, $a(E_2)$ and $a(E_1, E_2)$ into a vector z, and perform hyperbolic tangent $tanh$ to generate a hidden layer,

$$z = a(E_1) \oplus a(E_2) \oplus a(E_1, E_2) \tag{12}$$

$$h = \tanh(z) \tag{13}$$

where $z, h \in \mathrm{R}^{9\times(d+l)}$.

3.4 Relation Classification

At last, we apply the softmax function upon the hidden layer to predict K-class classification,

$$y = f(\mu h + b) \tag{14}$$

where K is the size of the event relation categories, and $\mu \in \mathrm{R}^{K\times 9\times(d+l)}$ is parameter matrix, $b \in \mathrm{R}^{9\times(d+l)}$ is a bias term. We compute the cross-entropy error between y and gold relation g, and further define the objective function:

$$J(\theta) = -\sum_{s=1}^{S}\sum_{k}^{K} y_k(s) \log g_k(s) + \frac{1}{2}\lambda||\theta||^2 \tag{15}$$

where s is the s-th instance in training set S, k is relation type in K, and $\theta = (w, f, l, \mu, b)$ is parameters to be learned.

Table 2. Distributions of positive and negative instances in training (Train), development (Dev) and test (Test) sets.

Data	Positive/Negative			
	Comparison	Contingency	Expansion	Temporal
Train	617/617	708/708	1520/1520	547/547
Dev	145/815	233/727	486/474	96/864
Test	197/878	263/812	514/561	101/974

4 Experiments

4.1 Datasets

We utilize 968 event pairs [14] annotated on FrameNet-1.5 [8], and follow their annotation metric to annotate 4459 new pairs on GIGAWORD (LDC2003T05), both of which finally make up of our experimental datasets in Table 2.

4.2 Experimental Setup

Word embedding is trained on large-scale data by word2vec toolkit. We empirically adopt the same parameters in our experiments. Specifically, we set d=200, l=5 and batch=128. We apply stochastic gradient descent (SGD) algorithm to minimize $J(\theta)$, with learning rate lr=0.1 and momentum=0.9. Finally, we choose precision (P), recall (R) and F_1-score (F_1) as evaluation metrics.

4.3 Comparison with state-of-the-art methods

Table 3. The experimental results of different models. EL=Event-Level and CE=Cross-Event.

Relations	**Models**	**Performance (%)**		
		P	**R**	$\mathbf{F_1}$
Comparison	Cross-Scenario	24.90	60.91	35.35
	Tri-Training	33.89	51.27	40.81
	SCNN(EL)	33.43	58.38	42.51
	SCNN(CE)	32.10	52.79	39.92
	SCNN(EL+CE)	34.04	56.85	42.59
	SCNN(EL+CE+Frame)	34.50	59.90	**43.78**
Contingency	Cross-Scenario	33.04	28.14	30.39
	Tri-Training	34.83	50.19	41.12
	SCNN(EL)	32.80	61.98	42.89
	SCNN(CE)	35.69	49.81	41.59
	SCNN(EL+CE)	34.42	60.46	43.86
	SCNN(EL+CE+Frame)	37.09	56.27	**44.71**
Expansion	Cross-Scenario	53.57	58.37	55.87
	Tri-Training	52.34	67.51	58.96
	SCNN(EL)	55.84	63.23	59.31
	SCNN(CE)	55.53	59.53	57.46
	SCNN(EL+CE)	56.13	67.70	**61.38**
	SCNN(EL+CE+Frame)	56.28	65.37	60.49
Temporal	Cross-Scenario	17.33	34.65	23.10
	Tri-Training	19.22	63.37	29.49
	SCNN(EL)	19.01	53.47	28.05
	SCNN(CE)	19.26	51.49	28.03
	SCNN(EL+CE)	20.62	52.48	29.61
	SCNN(EL+CE+Frame)	21.25	57.43	**31.02**

We select the following state-of-the-art methods for comparison:

Cross-Scenario: [14] propose a cross-scenario inference method to predict relations between pairwise events, which assumes that events with same scenarios share the similar relations.

Tri-Training: [6] propose a semi-supervised learning method based on Tri-Training, which improves the classification performance by expanding training corpus with higher confidence unlabelled samples.

SCNN (EL): [15] succeed in performing a SCNN into implicit discourse relation recognition, which could be also applied to ERC task. In this system, only Event-Level(EL) features are taken into account, we set $e_i = w_i$ in Eq. 3 and $z = a(E_1) \oplus a(E_2)$ in Eq. 12, $z, h \in \mathrm{R}^{6d}$.

SCNN (CE): For sufficient and valid comparison, we further set $e_i = w_i$ in Eq. 3 and $z = a(E_1, E_2)$ in Eq. 12, only take Cross-Event(CE) features into account, $z \in \mathrm{R}^{3d}$.

SCNN (EL + CE): This system combines event-level together with cross-event features without using frame embedding features. Specifically, we set $e_i = w_i$ in Eq. 3 and $z = a(E_1) \oplus a(E_2) \oplus a(E_1, E_2)$ as Eq. 12 exhibits, $z \in \mathrm{R}^{9d}$.

SCNN (EL + CE + Frame): At last, we combine event-level and cross-event features together with frame embedding features, we set $e_i = w_i \oplus f_i$ as Eq. 3 and $z = a(E_1) \oplus a(E_2) \oplus a(E_1, E_2)$ as Eq. 12 exhibits, $z \in \mathrm{R}^{9\times(d+l)}$.

4.4 Results and Analysis

As shown in Table 3, we found the model SCNN (EL) performs better than Cross-Scenario and Tri-Training, which suggests that the shallow structure works well and using event-level features is beneficial to ERC.

When looking into SCNN (CE), we observe that this method works worse than Tri-Training and SCNN (EL). The main reason may be that considering cross-event features only might have left out the important information in each event. However, SCNN (CE) performs better than Cross-Scenario in general, which reveals that the cross-event features is also effective to some extent.

Table 4. The overall performance of different models using macro average measure.

Models	**Macro-average(%)**		
	P	**R**	**$\mathbf{F_1}$**
Cross-Scenario	32.21	45.52	37.73
Tri-Training	35.07	58.09	43.74
SCNN(EL)	36.34	58.79	44.92
SCNN(CE)	35.65	53.41	42.75
SCNN(EL+CE)	36.30	59.37	45.05
SCNN(EL+CE+Frame)	37.28	59.74	**45.91**

Compared with previous models, SCNN (EL+CE) achieves the best results in four relations and gets highest F_1-score 44.71% in `Expansion` among all models, which sufficiently suggests that combining event-level and cross-event information in relation classification gains remarkable promotion.

When frame embedding is merged into word embedding, we achieve the best result in `Comparison`, `Contingency` and `Temporal` in SCNN (EL+CE+Frame), which indicates that frame embedding is useful to represent the deep semantics of event relation. Table 4 presents the overall performance of six models using macro average measure, SCNN (EL+CE+Frame) model also achieves the best result. In summary, according to Table 3 and Table 4, our model undoubtedly gains the best result in ERC.

5 Conclusion

In this paper, we exploit a novel method for event relation classification which automatically extracts event-level and cross-event convolutional features from combined embeddings. We further concatenate these convolutional features into shallow neural networks to learn better classifiers. Experimental results show that our model significantly outperforms the state-of-the-art methods.

References

1. Aharon, R.B., Szpektor, I., Dagan, I.: Generating entailment rules from framenet. In: Proceedings of the ACL 2010 Conference Short Papers. pp. 241–246. Association for Computational Linguistics (2010)
2. Baroni, M., Dinu, G., Kruszewski, G.: Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In: ACL (1). pp. 238–247 (2014)
3. Burchardt, A., Frank, A.: Approaching textual entailment with lfg and framenet frames. In: Proc. of the Second PASCAL RTE Challenge Workshop.[-]. Citeseer (2006)
4. Chklovski, T., Pantel, P.: Global path-based refinement of noisy graphs applied to verb semantics. In: Natural Language Processing–IJCNLP 2005, pp. 792–803. Springer (2005)
5. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. The Journal of Machine Learning Research 12, 2493–2537 (2011)
6. Ding, S., Hong, Y., Zhu, S., Yao, J., Zhu, Q.: Research of event relation classification based on tri-training. The Journal of Computer Engineering and Science 37(12), 2345–2351 (2015)
7. Fillmore, C.: Frame semantics. Linguistics in the morning calm pp. 111–137 (1982)
8. Fillmore, C.J., Johnson, C.R., Petruck, M.R.: Background to framenet. International journal of lexicography 16(3), 235–250 (2003)
9. Harris, Z.: Mathematical structures of language. (1968)
10. Lin, D., Pantel, P.: Dirt@ sbt@ discovery of inference rules from text. In: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 323–328. ACM (2001)
11. Pantel, P., Pennacchiotti, M.: Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. pp. 113–120. Association for Computational Linguistics (2006)

12. Shen, D., Lapata, M.: Using semantic roles to improve question answering. In: EMNLP-CoNLL. pp. 12–21 (2007)
13. Socher, R., Huang, E.H., Pennin, J., Manning, C.D., Ng, A.Y.: Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In: Advances in Neural Information Processing Systems. pp. 801–809 (2011)
14. Yang, X., Hong, Y., Chen, Y., Wang, X., Yao, J., Zhu, Q.: Detection event relation through cross-scenario inference. The Journal of Chinese Information Processing 28(5), 206–214 (2014)
15. Zhang, B., Su, J., Xiong, D., Lu, Y., Duan, H., Yao, J.: Shallow convolutional neural network for implicit discourse relation recognition. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 2230–2235 (2015)