

Learning Constraints and Descriptive Segmentation for Subevent Detection

Haoyu Wang¹, Hongming Zhang^{2,*}, Muhao Chen³ & Dan Roth¹

¹Department of Computer and Information Science, UPenn

²Department of Computer Science and Engineering, HKUST

³Department of Computer Science, USC

{why16gz1, danroth}@seas.upenn.edu;

hzhangal@cse.ust.hk; muhaoche@usc.edu

Abstract

Event mentions in text correspond to real-world events of varying degrees of granularity. The task of subevent detection aims to resolve this granularity issue, recognizing the membership of multi-granular events in event complexes. Since knowing the span of descriptive contexts of event complexes helps infer the membership of events, we propose the task of *event-based text segmentation* (EVENTSEG) as an auxiliary task to improve the learning for subevent detection. To bridge the two tasks together, we propose an approach to learning and enforcing constraints that capture dependencies between subevent detection and EVENTSEG prediction, as well as guiding the model to make globally consistent inference. Specifically, we adopt Rectifier Networks for constraint learning and then convert the learned constraints to a regularization term in the loss function of the neural model. Experimental results show that the proposed method outperforms baseline methods by 2.3% and 2.5% on benchmark datasets for subevent detection, HiEve and IC, respectively, while achieving a decent performance on EVENTSEG prediction¹.

1 Introduction

Since real-world events are frequently conveyed in human languages, understanding their linguistic counterparts, i.e. event mentions in text, is of vital importance to natural language understanding (NLU). One key challenge to understanding event mentions is that they refer to real-world events with varied granularity (Glavaš et al., 2014) and form *event complexes* (Wang et al., 2020). For example, when speaking of a coarse-grained event “publishing a paper”, it can involve a complex of

Former Penn State football coach Jerry Sandusky **posted** (e1) bail Thursday after spending a night in jail following a new round of sex-abuse **charges** (e2) filed against him. Sandusky secured his release **using** (e3) \$200,000 in real estate holdings and a \$50,000 certified check **provided** (e4) by his wife, Dorothy, according to online court record ... He was also **charged** (e5) last month with abusing eight boys, some on campus, over 15 years, allegations that were not immediately brought to the attention of authorities even though high-level people at Penn State apparently knew about them. In all, he faces more than 50 **charges** (e6). The **scandal** (e7) has resulted in the **ousting** (e8) of school President Graham Spanier and longtime coach Joe Paterno.

Figure 1: An example of PARENT-CHILD relations and EVENTSEGs from the HiEve dataset (Glavaš et al., 2014). The blue and yellow segments denote the textual spans of event complexes “posted” and “scandal” respectively. Curved arrows denote PARENT-CHILD relations within a text segment, whereas the dotted arrows denote cross-segment PARENT-CHILD relations.

more fine-grained events such as “writing the paper,” “passing the peer review,” and “presenting at the conference.” Naturally, understanding events requires to resolve the granularity of events and infer their memberships, which corresponds to the task of subevent detection (a.k.a. event hierarchy extraction). Practically, subevent detection is a key component of event-centric NLU (Chen et al., 2021), and is beneficial to various applications, such as schema induction (Zhang et al., 2020; Li et al., 2020a), task-oriented dialogue agents (Andreas et al., 2020), summarization (Chen et al., 2019; Zhao et al., 2020), and risk detection (Pohl et al., 2012).

As a significant step towards inducing event complexes (graphs that recognize the relationship of multi-granular events) in documents, subevent detection has started to receive attention recently (Wang et al., 2020; Han et al., 2021). It is natu-

* This work was done when the author was visiting the University of Pennsylvania.

¹Our code is publicly available at http://cogcomp.org/page/publication_view/950.

ral to perceive that in documents, there might be several different event complexes and they often span in different descriptive contexts that form relatively independent text segments. Consider the example in Fig. 1, where the two membership relations in the event complex (graph consisting of “scandal (e7),” “charges (e6),” “ousting (e8),” and relations) are both within the segment marked in yellow that describes the event complex. As can be seen in the paragraph, though we cannot deny the existence of cross-segment subevent relations (dotted arrows), events belong to the same membership are much more often to co-occur in a text segment. This correlation has been overlooked by existing data-driven methods (Zhou et al., 2020; Yao et al., 2020), which formulate subevent detection as pairwise relation extraction. On the other hand, while prior studies have demonstrated the benefits of incorporating logical constraints among event memberships and other relations (such as co-reference) (Glavaš and Šnajder, 2014; Wang et al., 2020), the constraints between the memberships and event co-occurrences in text segments remain uncertain. Hence, how to effectively learn and enforce hard-to-articulate constraints as in the case of subevent detection and segmentation of text is another challenge.

Our *first* contribution is to improve subevent detection based on an auxiliary task of EVENTSEG prediction. By EVENTSEG prediction, we seek to segment a document into descriptive contexts of different event complexes. Evidently, with EVENTSEG information, it would be relatively easy to infer the memberships of events in the same descriptive context. Using annotations for subevent detection and EVENTSEG prediction, we aim to adopt a neural model to jointly learn these two tasks along with the (soft) logical constraints that bridge their labels together. In this way, we incorporate linear discourse structure of segments into membership relation extraction, avoiding complicated feature engineering in the previous work (Aldawsari and Finlayson, 2019). From the learning perspective, adding EVENTSEG prediction as an auxiliary task seeks to provide effective incidental supervision signals (Roth, 2017) to the subevent detection task. This is especially important in the current scenario where annotated learning resources for subevents are typically limited (Hovy et al., 2013; Glavaš et al., 2014; O’Gorman et al., 2016).

To capture the logical dependency between

subevent structure and EVENTSEG, our *second* contribution is an approach to automatically learning and enforcing logical constraints. Motivated by Pan et al. (2020), we use Rectifier Networks to learn constraints in the form of linear inequalities, and then convert the constraints to a regularization term that can be incorporated into the loss function of the neural model. This allows any hard-to-articulate constraints to be automatically captured for inter-related tasks, and efficiently guides the model to make globally consistent inference. By learning and enforcing task-specific constraints for subevent relations, the proposed method achieves comparable results with SOTA subevent detection methods on the HiEve and IC dataset. Moreover, by jointly learning with EVENTSEG prediction, the proposed method surpasses previous methods on subevent detection by relatively 2.3% and 2.5% in F_1 on HiEve and IC, while achieving decent results on EVENTSEG prediction.

2 Related Work

We discuss three lines of relevant research.

Subevent Detection. Several approaches to extracting membership relations have been proposed, which mainly fall into two categories: statistical learning methods and data-driven methods. Statistical learning methods (Glavaš et al., 2014; Glavaš and Šnajder, 2014; Araki et al., 2014; Aldawsari and Finlayson, 2019) collect a variety of features before feeding into classifiers for pairwise decision. Nevertheless the features often require costly human effort to obtain, and are often dataset-specific. Data-driven methods, on the other hand, automatically characterize events with neural language models like BERT (Devlin et al., 2019), and can simultaneously incorporate various signals such as event time duration (Zhou et al., 2020), joint constraints with event temporal relations (Wang et al., 2020) and subevent knowledge (Yao et al., 2020). Among recent methods, only Aldawsari and Finlayson (2019) utilize discourse features like discourse relations between elementary discourse units, but still document-level segmentation signals are not incorporated into the task of subevent detection. Actually, research on event-centric NLU (Chen et al., 2021) has witnessed the usage of document-level discourse relations: different functional discourse structures around the main event in news articles have been studied in Choubey et al. (2020). Hence, we attempt to capture the interdependencies

between subevent detection and segmentation of text, in order to enhance the model performance for event hierarchy extraction.

Text Segmentation. Early studies in this line have concentrated on unsupervised text segmentation, quantifying lexical cohesion within small text segments (Choi, 2000), and unsupervised Bayesian approaches have also been successful in this task (Eisenstein and Barzilay, 2008; Eisenstein, 2009; Newman et al., 2012; Mota et al., 2019). Given that unsupervised algorithms are difficult to specialize for a particular domain, Koshorek et al. (2018) formulate the problem as a supervised learning task. Lukasik et al. (2020) follow this idea by using transformer-based architectures with cross segment attention to achieve state-of-the-art performance. Focusing on creating logically coherent sub-document units, these prior work do not cover segmentation of text regarding descriptive contexts of event complexes, which is the focus of the auxiliary task in this work.

Learning with Constraints. In terms of enforcing declarative constraints in neural models, early efforts (Roth and Yih, 2004; Glavaš and Šnajder, 2014) formulate the inference process as an Integer Linear Programming (ILP) problem. Pan et al. (2020) also employ ILP to enforce constraints learned automatically from Rectifier Networks with strong expressiveness (Pan and Srikumar, 2016). Yet the main drawback of solving an ILP problem is its inefficiency in a large feasible solution space. Recent work on integrating neural networks with structured outputs has emphasized the importance of the interaction between constraints and representations (Rocktäschel and Riedel, 2017; Niculae et al., 2018; Li and Srikumar, 2019; Li et al., 2019, 2020b). However there has been no automatic and efficient ways to learn and enforce constraints that are not limited to first-order logic, e.g., linear inequalities learned via Rectifier Networks. And this is the research focus of our paper.

3 Preliminaries

A document $\mathcal{D}$ consists of a collection of m sentences $\mathcal{D} = [s_1, s_2, \dots, s_m]$, and each sentence, say s_k , contains a sequence of tokens $s_k = [w_1, w_2, \dots, w_n]$. Some tokens in sentences belong to the set of annotated event triggers, i.e., $\mathcal{E}_{\mathcal{D}} = \{e_1, e_2, \dots, e_l\}$. Following the notation by Koshorek et al. (2018), a segmentation of doc-

ument $\mathcal{D}$ is represented as a sequence of binary values: $\mathcal{Q}_{\mathcal{D}} = \{q_1, q_2, \dots, q_{m-1}\}$, where q_i indicates whether sentence s_i is the end of a segment.

Subevent Detection is to identify membership relations between events, given event mentions in documents. Particularly, $\mathcal{R}$ denotes the set of relation labels as defined in Hovy et al. (2013) and Glavaš et al. (2014) (i.e., PARENT-CHILD, CHILD-PARENT, COREF, and NOREL). For a relation $r \in \mathcal{R}$, we use a binary indicator $Y_{i,j}^r$ to denote whether an event pair (e_i, e_j) has relation r , and use $y_{i,j}^r$ to denote the model-predicted possibility of an event pair (e_i, e_j) to have the relation r .

EventSeg prediction aims at finding an optimal segmentation of text that breaks the document into several groups of consecutive sentences, and each sequence is a descriptive context of an event complex (Wang et al., 2020). Being different from the traditional definition of text segmentation, EVENTSEG focuses on the change of event complex (which is not necessarily the change of topic). For a pair of events (e_i, e_j) , we use a binary indicator $Z_{i,j}$ to denote whether the two events are within the same descriptive context of event complexes, and $z_{i,j}$ to denote the model-predicted possibility of two events to belong to the same segment. Details on how to obtain EVENTSEG are described in §5.1.

Connections between Two Tasks. Statistically, through an analysis of the HiEve and IC corpus, PARENT-CHILD and CHILD-PARENT relations appear within the same descriptive context of event complex with a probability of 65.13% (see Tab. 1). On the other hand, the probability for each of the two other non-membership relations (i.e., COREF and NOREL) to appear within the same segment approximately equals that of its appearance across segments. This demonstrates that subevent relations tend to appear within the same EVENTSEG. Since this is not an absolute logical constraint, we adopt an automatic way of modeling such constraints instead of manually inducing them, which is described in the next section.

4 Methods

We now present the framework for learning and enforcing constraints for the main task of subevent detection and the auxiliary EVENTSEG prediction. We start with learning the hard-to-articulate constraints (§4.1), followed by details of joint learning (§4.2) and inference (§4.3) for the two tasks.

4.1 Learning Constraints

From the example shown in Fig. 1 we can construct an event graph G with all the events, membership relations, and EVENTSEG information. Fig. 2 shows a three-event subgraph of G . The goal of constraint learning is as follows: given membership relations $Y_{i,j}^r, Y_{j,k}^r$ and segmentation information $Z_{i,j}, Z_{j,k}$ about event pairs (e_i, e_j) and (e_j, e_k) , we would like to determine whether a certain assignment of $Y_{i,k}^r$, and $Z_{i,k}$ is legitimate.

Feature Space for Constraints. We now define the feature space for constraint learning. Specifically, we construct three-event subgraphs from documents, and a binary label t for structure legitimacy is created for each subgraph with membership relations and EVENTSEG information. Note that the feature space of Y^r and Z is not sufficient for capturing the dependency between two tasks. For instance, if $r = \text{PARENT-CHILD}$, $Y_{i,j}^r = Y_{j,k}^r = 1$, and $Z_{i,j} = Z_{j,k} = 0$, then due to the transitivity of PARENT-CHILD relation, we should enforce $Y_{i,k}^r = 1$. However, we cannot tell whether e_i and e_k are in the same EVENTSEG, i.e., both $Z_{i,k} = 1$ and $Z_{i,k} = 0$ could be legitimate. In other words, the current feature space is not separable for Z , and thus we need to expand the constraint features to better capture relationship legitimacy that involves three pairs of events. Let $\mathbf{X}_p = \{Y_p^r, r \in \mathcal{R}\} \cup \{Z_p\}$ denote the original set of features for an event pair p . Given features $\mathbf{X}_{i,j}$ and $\mathbf{X}_{j,k}$, we would like to *determine the set of possible values* for $\mathbf{X}_{i,k}$. We employ the *power set* of $\mathbf{X}_{i,k}$, $\mathcal{P}(\mathbf{X}_{i,k})$, as our new features for event pair (e_i, e_k) . And now a subgraph with three events e_i, e_j , and e_k can be featurized as

$$\mathbf{X} = \mathbf{X}_{i,j} \cup \mathbf{X}_{j,k} \cup \mathcal{P}(\mathbf{X}_{i,k}). \quad (1)$$

Constraint Learning with Rectifier Network. Inspired by how constraints are learned for several structured prediction tasks (Pan et al., 2020), we represent constraints for a given subgraph-label pair $(\mathbf{X}, t)$ as K linear inequalities.² Formally, for an input $\mathbf{X}$, we say the output t is feasible if it satisfies constraints c_k for all $k = 1, \dots, K$. And the k^{th} constraint c_k is expressed by a linear inequality

$$\mathbf{w}_k \cdot \mathbf{X} + b_k \geq 0,$$

²Here we assume K constraints is the upper bound for all the rules to be learned.

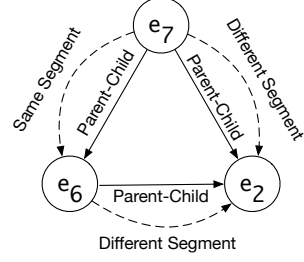


Figure 2: A legitimate structure for three-event subgraph obtained from the example shown in Fig. 1. The constraint features for the subgraph can be expressed by $\mathbf{X} = \mathbf{X}_{7,6} \cup \mathbf{X}_{6,2} \cup \mathcal{P}(\mathbf{X}_{7,2})$, and the label t for this structure is 1. The total number of features for constraint learning is $|\mathbf{X}| = 5 + 5 + 2^5 = 42$.

whose weights $\mathbf{w}_k$ and bias b_k are learned. Since a system of linear inequalities is proved to be equivalent to the Rectifier Network proposed in Pan et al. (2020), we adopt a two-layer rectifier network for learning constraints

$$p = \sigma \left(1 - \sum_{k=1}^K \text{ReLU}(\mathbf{w}_k \cdot \mathbf{X} + b_k) \right), \quad (2)$$

where p denotes the possibility of $t = 1$ and $\sigma(\cdot)$ denotes the sigmoid function. We train the parameters $\mathbf{w}_k$'s and b_k 's of the rectifier network in a supervised setting. The positive examples are induced from subgraph structures that appear in the training corpus, while the negative examples are randomly chosen from the rest possibilities that do not exist in the training corpus.

4.2 Joint Task Learning

After learning the constraints using Rectifier Networks, we introduce how to jointly model membership relations and EVENTSEG with neural networks and how to integrate the learned constraints into the model. The model architecture is shown in Fig. 3.

Local Classifier. To characterize event pairs in documents, we employ a neural encoder, which obtains contextualized representations for event triggers from the pre-trained transformer-based language model RoBERTa (Liu et al., 2019). As the context of event pairs, the sentences where two event mentions appear are concatenated using [CLS] and [SEP]. We then calculate the element-wise average of subword-level contextual representations as the representation for each event trigger. To obtain event pair representation for (e_i, e_j) , we concatenate the two contextual representations, together with their element-wise Hadamard product

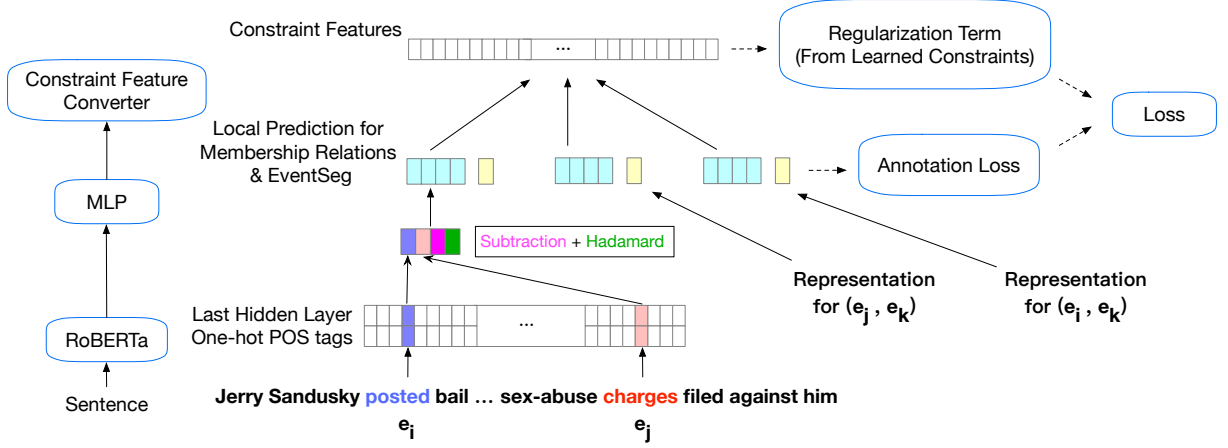


Figure 3: An overview of our approach. The model takes three pairs of events at a time in training to enforce constraints over three-event subgraphs (an example can be found in Fig. 2). Event pair representations are obtained from RoBERTa where the context between two events are taken into consideration. Soft logical constraints learned in §4.1 are converted to a regularization term in the loss function for subgraph structure legitimacy.

and subtraction as in Wang et al. (2020). The event pair representation is then sent to a multi-layer perceptron (MLP) with $|\mathcal{R}|$ outputs for estimation of the confidence score $y_{i,j}^r$ for each relation r . To make EVENTSEG as an auxiliary task, the model also predicts whether two events belong to the same segment using another separate MLP with a single-value output $z_{i,j}$. In accordance with the learned constraints in §4.1, the model takes three pairs of events at a time. The annotation loss in Fig. 3 is a linear combination of a four-class cross-entropy loss $L_{A,sub}$ for subevent detection and a binary cross-entropy loss $L_{A,seg}$ for EVENTSEG.

Incorporating Subgraph Constraints. The K constraints learned in §4.1 are encoded into the weights $\mathbf{w}_k$ and bias b_k , $k = 1, \dots, K$. Now that the input $\mathbf{X}$ is considered valid if it satisfies all K constraints, we obtain the predicted probability p of $\mathbf{X}$ being valid from Eq. 2. To add the constraints as a regularization term in the loss function of the neural model, we convert p into the negative log space (Li et al., 2019) which is same as the cross-entropy loss. And thus the loss corresponding to the learned constraints is

$$L_{cons} = -\log\left(\text{Sigmoid}\left(1 - \sum_{k=1}^N \text{ReLU}(\mathbf{w}_k \cdot \boldsymbol{\psi} + b_k)\right)\right).$$

And the loss function of the neural model is

$$L = \lambda_1 L_{A,sub} + \lambda_2 L_{A,seg} + \lambda_3 L_{cons}, \quad (3)$$

where the λ 's are non-negative coefficients to control the influence of each loss term. With the loss function in Eq. 3, we train the model in a supervised way to fine-tune RoBERTa.

4.3 Inference

At inference time, to extract relations in the subevent detection task, we input a pair of events into the model and compare the predicted possibility for each relation, leaving the other two input pairs blank. For EVENTSEG prediction, we let the model predict $z_{i,i+1}$ for each pair of adjacent events (e_i, e_{i+1}) that appear in different sentences. If $z_{i,i+1} = 1$, it means there is a segment break between e_i and e_{i+1} . When there are intermediate sentences between the two adjacent event mentions, we treat the sentence that contains e_i as the end of a previous segment. In this way, we provide an approach to solving two tasks together via automatically learning and enforcing constraints in the neural model. We provide in-depth experimentation for the proposed method in the next section.

5 Experiments

Here we describe the experiments on subevent detection with EVENTSEG prediction as an auxiliary task. We first introduce the corpora used (§5.1), followed by evaluation for subevent detection and an ablation study for illustrating the importance of each model component (§5.2-§5.4). We also provide a case study on EVENTSEG prediction (§5.5) and an analysis of the constraints learned in the model (§5.6).

5.1 Datasets

HiEve The HiEve corpus (Glavaš et al., 2014) contains 100 news articles. Within each article, annotations are given for both subevent membership

Relations	HiEve		IC	
	Within	Across	Within	Across
Parent-Child	1,123	679	1,698	550
Child-Parent	1,067	779	1,475	863
Coref	322	436	1,476	877
NoRel	32,029	31,726	40,072	41,815

Table 1: Statistics of the HiEve and IC dataset. Numbers in column “Within” denote the number of relations appearing within the same descriptive context of event complex, whereas numbers under “Across” denote those across different segments.

and coreference relations. Using the same measurement of inter-annotator agreement (IAA) as event temporal relations in [UzZaman and Allen \(2011\)](#), the HiEve dataset has an IAA of 0.69 F1.

Intelligence Community (IC) The IC corpus ([Hovy et al., 2013](#)) also contains 100 news articles annotated with membership relations. The articles report violence events such as attack, war, etc. We discard those relations involving implicit events annotated in IC, and calculate transitive closure for both subevent relations and co-reference to get annotations for all event pairs in text order as it is done for HiEve ([Glavaš et al., 2014](#)).

Labeling EVENTSEG We explain how to segment the document using annotations for subevent relations. First, we use the annotated subevent relations (PARENT-CHILD and CHILD-PARENT only) to construct a directed acyclic event graph for each document. Due to the property of subevent relations, each connected component in the graph is actually a tree with one root node, which forms an event complex. If the graph constructed from document has one connected component, we remove the root node and separate the event graph into more than one event complexes. Since each event complex has a textual span in the document, we obtain several descriptive contexts that may or may not overlap with each other. For those documents with non-overlapping descriptive contexts, their segmentations are therefore obtained. In cases where two descriptive contexts of event complexes overlap with each other, if there exists such an event whose removal results in non-overlapping contexts, then we segment the contexts assuming this event is not considered. Otherwise, we merge the contexts into one segment. Through this event-based text segmentation, on average we obtain 3.99 and 4.29 EVENTSEGs in the HiEve and IC corpus,

respectively.

We summarize the data statistics in [Tab. 1](#).

5.2 Baselines and Evaluation Protocols

On IC dataset, we compare with two baseline approaches. [Araki et al. \(2014\)](#) propose a logistic regression model along with a voting algorithm for parent event detection. [Wang et al. \(2020\)](#) use a data-driven model that incorporates hand-crafted constraints with event temporal attributes to extract event-event relations. On Hieve³, we compare with a transformer-based language model TACOLM ([Zhou et al., 2020](#)) that fine-tunes on a temporal common sense corpora, and the method proposed by [Wang et al. \(2020\)](#) which also serves as the second baseline for IC.

We use the same evaluation metric on HiEve as previous methods ([Zhou et al., 2020](#)), leaving 20% of the documents out for testing⁴. The F_1 scores of PARENT-CHILD and CHILD-PARENT and the micro-average of them are reported. In accordance with HiEve, the IC dataset is also evaluated with F_1 scores of membership relations instead of BLANC ([Araki et al., 2014](#)), while the other settings remain the same with previous works.

5.3 Experimental Setup

We fine-tune the pre-trained 1024 dimensional RoBERTa ([Liu et al., 2019](#)) to obtain contextual representations of event triggers in a supervised way given labels for membership relations and EVENTSEG. Additionally, we employ 18 dimensional one-hot vectors for part-of-speech tags for tokens in documents to include explicit syntactic features in the model. For each MLP we set the dimension to the average of the input and output neurons, following [Chen et al. \(2018\)](#). The parameters of the model are optimized using AMSGrad ([Reddi et al., 2018](#)), with the learning rate set to 10^{-6} . The training process is limited to 40 epochs since it is sufficient for convergence.

5.4 Results

We report the results for subevent detection on two benchmark datasets, HiEve and IC, in [Tab. 2](#).

³Despite carefully following the details described in [Al-dawsari and Finlayson \(2019\)](#) and communicating with the authors, we were not able to reproduce their results. Therefore, we choose to compare with other methods.

⁴To make predictions on event complexes, we keep all negative NOREL instances in our experiments instead of strictly following [Zhou et al. \(2020\)](#) and [Wang et al. \(2020\)](#) where negative instances are down-sampled with a probability of 0.4.

Corpus	Model	F_1 score		
		PC	CP	Avg.
IC	Araki et al. (2014)	-	-	0.262
	Wang et al. (2020)	0.421	0.495	0.458
	Our model	0.446	0.516	0.481
HiEve	Zhou et al. (2020)	0.485	0.494	0.489
	Wang et al. (2020)	0.472	0.524	0.497
	Our model	0.534	0.510	0.522

Table 2: Experimental results for subevent detection on IC and HiEve corpus. PC, CP and Avg. denote PARENT-CHILD, CHILD-PARENT and their micro-average, respectively. F_1 scores for PC and CP are not reported in Araki et al. (2014).

Among the baseline methods, Wang et al. (2020) has the best results in terms of F_1 on both datasets. They integrate event temporal relation extraction, common sense knowledge and handcrafted logical constraints into their approach. In contrast, our proposed method does not require constraints induced by domain experts, but still outperforms their F_1 score by 2.3 - 2.5%. We attribute this superiority to the use of connections between subevent relations and the linear discourse structure of segments. Thanks to the strong expressiveness of Rectifier Networks, we utilize these connections via the learning of linear constraints, thus incorporating incidental supervision signal from EVENTSEG. Furthermore, the event pair representation in our model is obtained from broader contexts than the local sentence-level contexts for events in Wang et al. (2020). The new representation not only contains more information on events but naturally provides necessary clues for determining whether there is a break for EVENTSEG.

We further perform an ablation analysis to aid the understanding of the model components and report our findings in Tab. 3. Without any constraints, integrating EVENTSEG prediction as an auxiliary task brings along an absolute gain of 0.2% and 0.6% in F_1 on HiEve and IC respectively over the vanilla single-task model with RoBERTa fine-tuning. This indicates that EVENTSEG information is beneficial to the extraction of membership relations. When membership constraints are added via the regularization term into the loss function, the model’s performance on subevent detection is significantly improved by 2.1% in F_1 on HiEve dataset. Incorporating constraints involving two tasks further enhances the model performance by 0.5% - 1.1%. This indicates that the global con-

sistency ensured within and across EVENTSEGs is important for enhancing the comprehension for subevent memberships.

5.5 Case Study for EVENTSEG Prediction

Here we provide an analysis of model performance on the task of EVENTSEG prediction. Though EVENTSEG prediction is somewhat different from text segmentation in concept, we can use methods for text segmentation as baselines for EVENTSEG prediction. We train a recent BERT-based model (Lukasik et al., 2020) for text segmentation based on annotations for EVENTSEG in the HiEve and IC corpora and compare our method with this baseline. In Tab. 4 we show the performances of the baseline model and ours for EVENTSEG prediction in terms of F_1 on HiEve and IC. Since our solution for EVENTSEG prediction is essentially similar to the cross-segment BERT model in terms of representations of segments, our performance is on par with the baseline model.

5.6 Analysis on Constraint Learning

We further provide an in-depth qualitative analysis on different types of logical constraints captured by the constraint learning.

5.6.1 Types of Learned Constraints

We expect that both task-specific constraints (membership relations only) in previous works (Glavaš and Šnajder, 2014; Wang et al., 2020) and cross-task constraints can be automatically captured in our framework. Accordingly, we separately analyze these two constraints.

Task-specific Constraints. Since we are using three-event subgraph for constraint learning, apparently, transitivity constraints for membership relations like

$$Y_{i,j}^r + Y_{j,k}^r - Y_{i,k}^r \leq 1,$$

$$r \in \{\text{PARENT-CHILD, CHILD-PARENT, COREF}\},$$

can be learned; whereas constraints that typically involve two events, e.g., symmetry constraints for membership relations like

$$Y_{i,j}^r = Y_{j,i}^{\bar{r}},$$

$$r \in \{\text{PARENT-CHILD, CHILD-PARENT}\},$$

can also be learned by assigning the third event e_k to the same event as e_i and treating the relation of (e_i, e_k) as COREF.

Model	HiEve			IC		
	P	R	F_1	P	R	F_1
Single-task Training	43.9	56.6	49.4	44.5	46.9	45.8
Joint Training	45.7	54.2	49.6	39.9	56.5	46.4
+ Membership Constraints	55.6	48.5	51.7	50.1	45.8	47.0
+ Membership + EVENTSEG	51.9	53.6	52.2	39.6	64.0	48.1

Table 3: Ablation study results for subevent detection. The results on both datasets are the micro-average of PARENT-CHILD and CHILD-PARENT in terms of precision, recall, and F_1 . “+ Membership Constraints” denotes adding automatically learned constraints for membership relations upon the joint training model. The row of “+ Membership + EVENTSEG” shows the results of the complete model.

Model	HiEve	IC
Cross-segment BERT (Lukasik et al., 2020)	55.2	58.3
Our model	56.8	57.4

Table 4: EVENTSEG prediction performance in terms of F_1 on the HiEve and IC corpus.

Cross-task Constraints. Here we provide an analysis of cross-task constraints for both membership relations and EVENTSEG information learned in the model. We give an example constraint in the form of linear inequality learned from HiEve

$$0.13x_0 + 0.19x_1 + 0.27x_2 + 0.08x_3 - 0.18x_4 + 0.09x_5 + 0.13x_6 + 0.25x_7 + 0.04x_8 - 0.18x_9 + \dots + 0.02x_{18} + 0.07x_{19} + \dots + 0.05 \geq 0,$$

where x_1 and x_6 denote the variables for $Y_{i,j}^r = 1$ and $Y_{j,k}^r = 1$ ($r = \text{CHILD-PARENT}$) respectively, and they both have positive coefficients. If we look at expected labels for $\mathcal{P}(\mathbf{X}_{i,k})$, we can see that x_{18} and x_{19} which denote the variables for $Y_{i,k}^r = 1, Z_{i,k} = 0$ and $Y_{i,k}^r = 1, Z_{i,k} = 1$ have coefficients of 0.02 and 0.07, respectively. The two positive coefficients for x_{18} and x_{19} indicate that (a) (e_i, e_k) is possible to have CHILD-PARENT relation, and (b) the possibility of (e_i, e_k) being in the same EVENTSEG is greater than two events being in different EVENTSEGs.

5.6.2 Qualitative Analysis

We set K to 10 since we observe less number of constraints will decrease the performance of learning accuracy while increasing K does not cause noticeable influence. We optimize the parameters using Adam with a learning rate of 0.001 and the training process is limited to 1,000 epochs. We show the performance of constraint learning in Tab. 5. Since the constraints for membership relations should be declarative hard constraints like symmetry and transitivity constraints in §5.6.1, the accuracy of constraint learning is equal or close to

Constraints	HiEve	IC
Membership	99.13	100.00
Membership + EVENTSEG	96.44	98.01

Table 5: Constraint learning performance in terms of accuracy on test set. “Membership” denotes the constraints involving membership relations only, while “Membership + EVENTSEG” denotes full constraints.

100%. Yet, those hard-to-articulate constraints that incorporate EVENTSEG information are more difficult to learn, and thus the Rectifier Network has a less satisfying performance in terms of accuracy on the test set of HiEve and IC (96.44% and 98.01%).

6 Conclusion

In this work we propose an automatic and efficient way of learning and enforcing constraints for subevent detection. By noticing the connections between subevent detection and EVENTSEG, we adopt EVENTSEG prediction as an auxiliary task which provides effective incidental supervision signals. Through learning and enforcing constraints that can express hard-to-articulate constraints, the logical rules for both tasks are captured to regularize the model towards consistent inference. The proposed approach outperforms SOTA data-driven methods on benchmark datasets and provides comparable results with recent text segmentation methods on EVENTSEG prediction. This demonstrates the effectiveness of the framework on subevent detection and the potential of solving other structured predictions tasks in NLP.

Ethical Considerations

This work does not present any direct societal consequence. The proposed method aims at supporting high-quality extraction of event complexes from documents with the awareness of discourse structures and automated constraint learning. We believe this study leads to intellectual merits of de-

veloping robust event-centric information extraction technologies. It also has broad impacts, since constraints and dependencies can be broadly investigated for label structures in various natural language classification tasks. The acquired eventually knowledge, on the other hand, can potentially benefit various downstream NLU and NLG tasks.

For any information extraction methods, real-world open source articles to extract information from may include societal biases. Extracting event complexes from articles with such biases may potentially propagate the bias into acquired knowledge representation. While not specifically addressed in this work, the ability to incorporate logical constraints and discourse consistency can be a way to mitigate societal biases.

Acknowledgement

We appreciate the anonymous reviewers for their insightful comments.

This research is supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2019-19051600006 under the BETTER Program, and by contract FA8750-19-2-1004 with the US Defense Advanced Research Projects Agency (DARPA). The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

References

- Mohammed Aldawsari and Mark Finlayson. 2019. [Detecting subevents using discourse and narrative features](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4780–4790, Florence, Italy. Association for Computational Linguistics.
- Jacob Andreas, John Bufe, David Burkett, Charles Chen, Josh Clausman, Jean Crawford, Kate Crim, Jordan DeLoach, Leah Dorner, Jason Eisner, et al. 2020. Task-oriented dialogue as dataflow synthesis. *Transactions of the Association for Computational Linguistics*, 8:556–571.
- Jun Araki, Zhengzhong Liu, Eduard Hovy, and Teruko Mitamura. 2014. [Detecting subevent structure for event coreference resolution](#). In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Muhao Chen, Changping Meng, Gang Huang, and Carlo Zaniolo. 2018. Neural article pair modeling for wikipedia sub-article matching. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 3–19. Springer.
- Muhao Chen, Hongming Zhang, Qiang Ning, Manling Li, Heng Ji, Kathleen McKeown, and Dan Roth. 2021. Event-centric natural language understanding. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics – Tutorials*.
- Xiuying Chen, Zhangming Chan, Shen Gao, Meng-Hsuan Yu, Dongyan Zhao, and Rui Yan. 2019. [Learning towards abstractive timeline summarization](#). In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 4939–4945. International Conferences on Artificial Intelligence Organization.
- Freddy Y. Y. Choi. 2000. [Advances in domain independent linear text segmentation](#). In *1st Meeting of the North American Chapter of the Association for Computational Linguistics*.
- Prafulla Kumar Choubey, Aaron Lee, Ruihong Huang, and Lu Wang. 2020. [Discourse as a function of event: Profiling discourse structure in news articles around the main event](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5374–5386, Online. 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. Association for Computational Linguistics.
- Jacob Eisenstein. 2009. [Hierarchical text segmentation from multi-scale lexical cohesion](#). In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 353–361, Boulder, Colorado. Association for Computational Linguistics.
- Jacob Eisenstein and Regina Barzilay. 2008. [Bayesian unsupervised topic segmentation](#). In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 334–343, Honolulu, Hawaii. Association for Computational Linguistics.
- Goran Glavaš and Jan Šnajder. 2014. [Constructing coherent event hierarchies from news stories](#). In *Proceedings of TextGraphs-9: the workshop on Graph-based Methods for Natural Language Processing*, pages 34–38, Doha, Qatar. Association for Computational Linguistics.
- Goran Glavaš, Jan Šnajder, Marie-Francine Moens, and Parisa Kordjamshidi. 2014. [HiEve: A corpus for](#)

- extracting event hierarchies from news stories. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3678–3683, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Rujun Han, I Hsu, Jiao Sun, Julia Baylon, Qiang Ning, Dan Roth, Nanyun Pen, et al. 2021. Ester: A machine reading comprehension dataset for event semantic relation reasoning. *arXiv preprint arXiv:2104.08350*.
- Eduard Hovy, Teruko Mitamura, Felisa Verdejo, Jun Araki, and Andrew Philpot. 2013. Events are not simple: Identity, non-identity, and quasi-identity. In *Workshop on Events: Definition, Detection, Coreference, and Representation*, pages 21–28, Atlanta, Georgia. Association for Computational Linguistics.
- Omri Koshorek, Adir Cohen, Noam Mor, Michael Rotman, and Jonathan Berant. 2018. Text segmentation as a supervised learning task. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 469–473, New Orleans, Louisiana. Association for Computational Linguistics.
- Manling Li, Qi Zeng, Ying Lin, Kyunghyun Cho, Heng Ji, Jonathan May, Nathanael Chambers, and Clare Voss. 2020a. Connecting the dots: Event graph schema induction with path language modeling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 684–695, Online. Association for Computational Linguistics.
- Tao Li, Vivek Gupta, Maitrey Mehta, and Vivek Srikumar. 2019. A logic-driven framework for consistency of neural models. 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 3924–3935, Hong Kong, China. Association for Computational Linguistics.
- Tao Li, Parth Anand Jawale, Martha Palmer, and Vivek Srikumar. 2020b. Structured tuning for semantic role labeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8402–8412, Online. Association for Computational Linguistics.
- Tao Li and Vivek Srikumar. 2019. Augmenting neural networks with first-order logic. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 292–302, Florence, Italy. Association for Computational Linguistics.
- 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.
- Michal Lukasik, Boris Dadachev, Kishore Papineni, and Gonalo Simões. 2020. Text segmentation by cross segment attention. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4707–4716, Online. Association for Computational Linguistics.
- Pedro Mota, Maxine Eskenazi, and Luísa Coheur. 2019. BeamSeg: A joint model for multi-document segmentation and topic identification. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 582–592, Hong Kong, China. Association for Computational Linguistics.
- David Newman, Nagendra Koilada, Jey Han Lau, and Timothy Baldwin. 2012. Bayesian text segmentation for index term identification and keyphrase extraction. In *Proceedings of COLING 2012*, pages 2077–2092, Mumbai, India. The COLING 2012 Organizing Committee.
- Vlad Niculae, Andre Martins, Mathieu Blondel, and Claire Cardie. 2018. SparseMAP: Differentiable sparse structured inference. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3799–3808. PMLR.
- Tim O’Gorman, Kristin Wright-Bettner, and Martha Palmer. 2016. Richer event description: Integrating event coreference with temporal, causal and bridging annotation. In *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*, pages 47–56, Austin, Texas. Association for Computational Linguistics.
- Xingyuan Pan, Maitrey Mehta, and Vivek Srikumar. 2020. Learning constraints for structured prediction using rectifier networks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4843–4858, Online. Association for Computational Linguistics.
- Xingyuan Pan and Vivek Srikumar. 2016. Expressiveness of rectifier networks. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2427–2435, New York, New York, USA. PMLR.
- Daniela Pohl, Abdelhamid Bouchachia, and Hermann Hellwagner. 2012. Automatic sub-event detection in emergency management using social media. In *Proceedings of the 21st international conference on world wide web*, pages 683–686.
- Sashank J Reddi, Satyen Kale, and Sanjiv Kumar. 2018. On the convergence of adam and beyond. In *International Conference on Learning Representations (ICLR)*.
- Tim Rocktäschel and Sebastian Riedel. 2017. End-to-end differentiable proving. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

- Dan Roth. 2017. [Incidental Supervision: Moving beyond Supervised Learning](#). In *Proc. of the Conference on Artificial Intelligence (AAAI)*.
- Dan Roth and Scott Yih. 2004. [A Linear Programming Formulation for Global Inference in Natural Language Tasks](#). In *Proc. of the Conference on Computational Natural Language Learning (CoNLL)*, pages 1–8. Association for Computational Linguistics.
- Naushad UzZaman and James Allen. 2011. [Temporal evaluation](#). In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 351–356, Portland, Oregon, USA. Association for Computational Linguistics.
- Haoyu Wang, Muhao Chen, Hongming Zhang, and Dan Roth. 2020. [Joint constrained learning for event-event relation extraction](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 696–706, Online. Association for Computational Linguistics.
- Wenlin Yao, Zeyu Dai, Maitreyi Ramaswamy, Bonan Min, and Ruihong Huang. 2020. [Weakly Supervised Subevent Knowledge Acquisition](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5345–5356, Online. Association for Computational Linguistics.
- Hongming Zhang, Muhao Chen, Haoyu Wang, Yangqiu Song, and Dan Roth. 2020. [Analogous process structure induction for sub-event sequence prediction](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1541–1550, Online. Association for Computational Linguistics.
- Lulu Zhao, Weiran Xu, and Jun Guo. 2020. [Improving abstractive dialogue summarization with graph structures and topic words](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 437–449, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Ben Zhou, Qiang Ning, Daniel Khashabi, and Dan Roth. 2020. [Temporal common sense acquisition with minimal supervision](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7579–7589, Online. Association for Computational Linguistics.