A KNOWLEDGE/DATA ENHANCED METHOD FOR JOINT EVENT AND TEMPORAL RELATION EXTRACTION

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

Xiaobin Zhang, Liangjun Zang,Peng Cheng,Yuqi Wang,Songlin Hu

Institute of Information Engineering, Chinese Academy of Sciences

School of Cyber Security, University of Chinese Academy of Sciences

State Key Laboratory of Media Convergence Production Technology and Systems

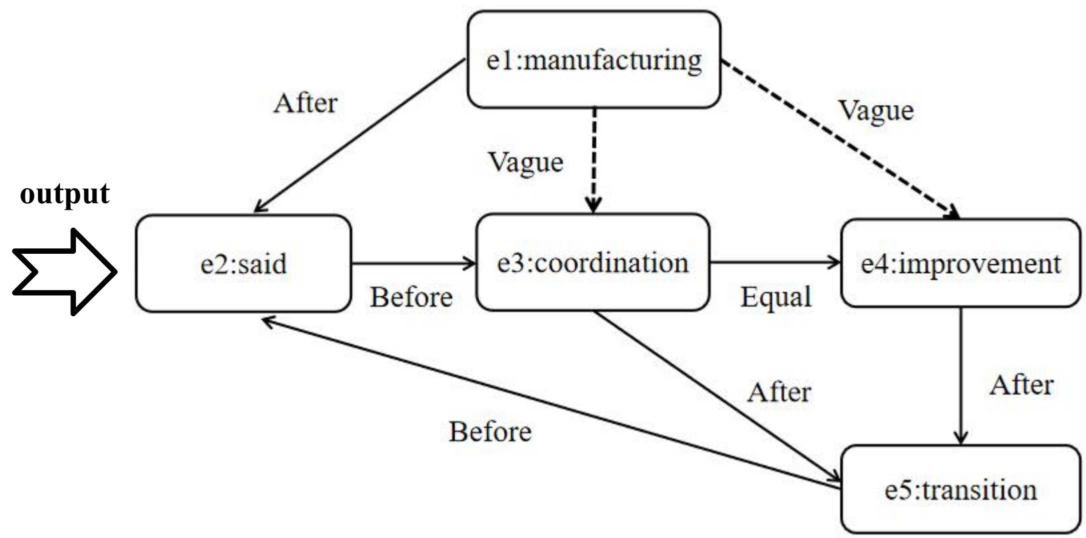
IEEE ICASSP 2022

Outline

Problem Definition

Given a raw text as input, our goal is to identifies event trigger words and classifies temporal relations for all predicted event pairs.

A major project is joint **(e1:manufacturing)** of an-70 cargo planes , the kremlin statement **(e2:said)** . the program also calls for **(e3:coordination)** of economic reforms and joint **(e4:improvement)** of social programs in the two countries , where many people have become impoverished during the chaotic post - soviet **(e5:transition)** to capitalism



Raw Corpus

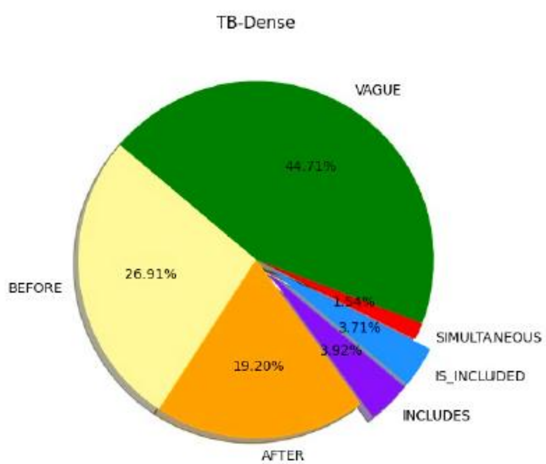
Event Trigger words and TempRel Relation

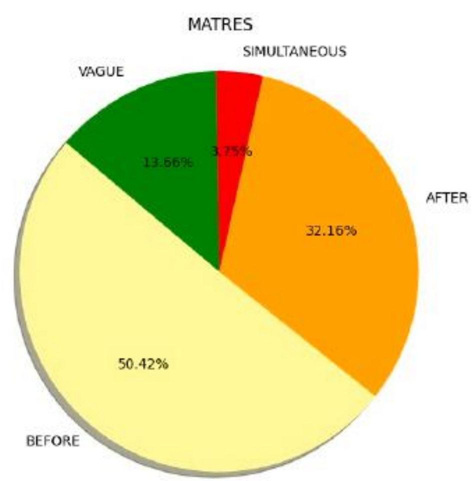
Challenges

- Limited amount of high-quality training data

Data Info	#of Documents			#of Pairs		
	Train	Dev	Test	Train	Dev	Test
TB-Dense	22	5	9	4032	629	1427
MATRES	183	-	20	6332	-	827

- Label imbalance

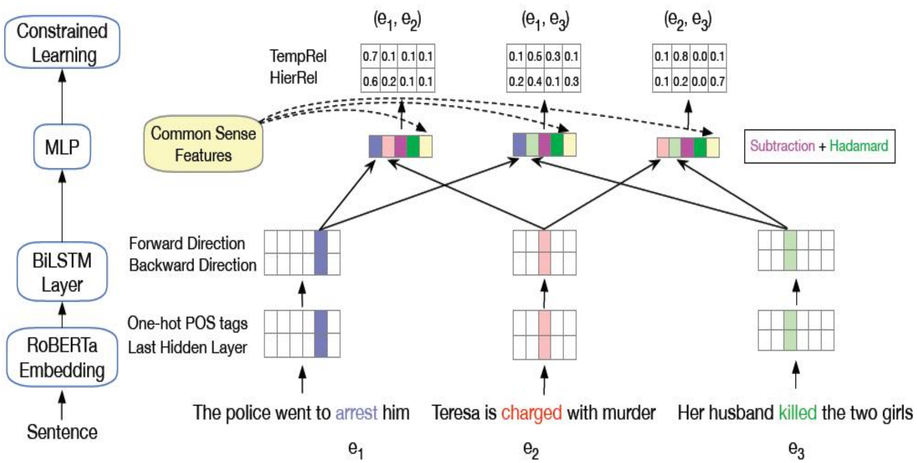




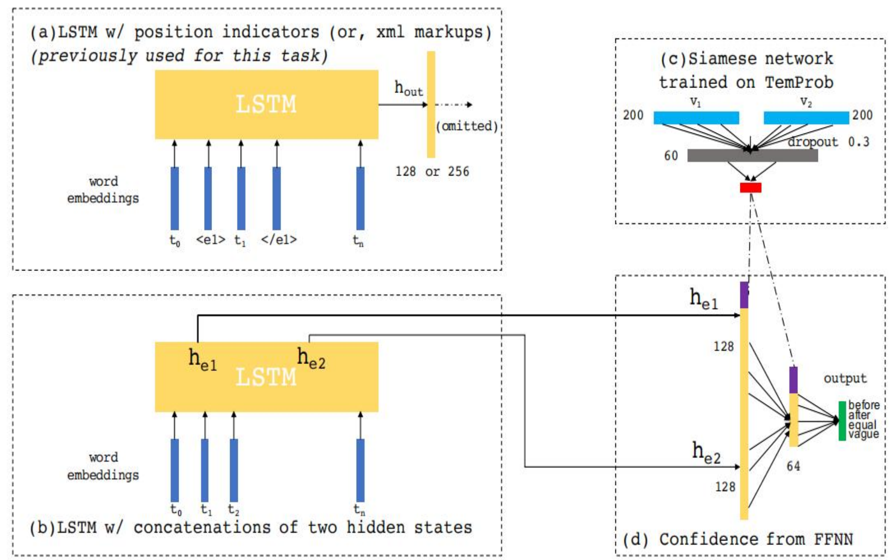
Related Work

Wang et al.(2020) adopt common sense features and a comprehensive set of logical constraints to extract TempRel and subevent relation.

Ning et al.(2019) use a Siamese network combined with trained by a knowledge base,global inference via ILP to extract TempRel relation.



$$L_C = \sum_{\substack{e_1,e_2,e_3 \in \mathcal{E}_D,\\ \alpha,\beta \in \mathcal{R}, \gamma \in \mathrm{De}(\alpha,\beta)}} |L_{t_1}| + \sum_{\substack{e_1,e_2,e_3 \in \mathcal{E}_D,\\ \alpha,\beta \in \mathcal{R}, \delta \notin \mathrm{De}(\alpha,\beta)}} |L_{t_2}|, \quad \begin{aligned} L_{t_1} &= \log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log \gamma_{(e_1,e_3)} \\ L_{t_2} &= \log \alpha_{(e_1,e_2)} + \log \beta_{(e_2,e_3)} - \log(1 - \delta_{(e_1,e_3)}) \end{aligned}$$

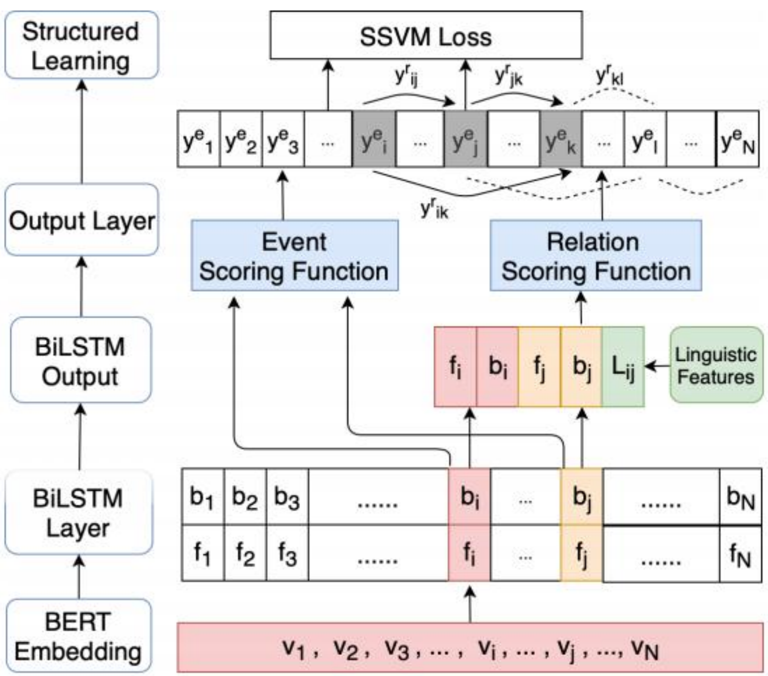


Wang et al.,Joint Constrained Learning for Event-Event Relation Extraction. EMNLP 2020.
Ning et al.,An Improved Neural Baseline for Temporal Relation Extraction. EMNLP 2019.

Related Work

Han et al.(2019) proposed a joint event and TempRel method with shared representation and structured prediction.

Wen et al.,(2021) proposed a schema-guided cross-document cross-lingual cross-media model for event and event relation extraction.



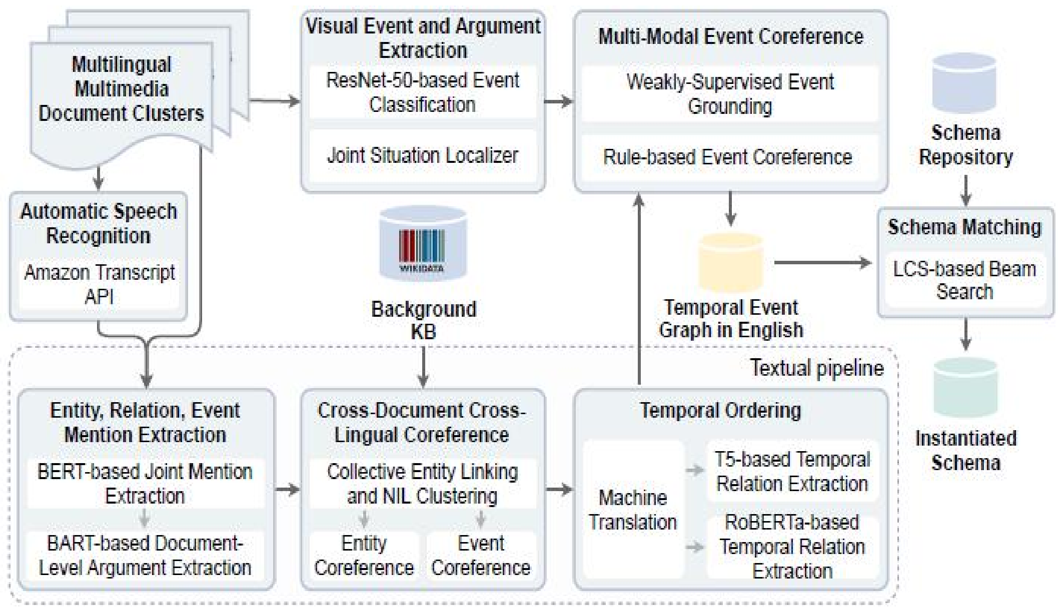


Figure 1: The architecture of RESIN schema-guided information extraction and temporal event tracking system.

Han et al.,Joint Event and Temporal Relation Extraction with Shared Representations and Structured Prediction. EMNLP 2019.
Wen et al.,RESIN: A Dockerized Schema-Guided Cross-document Cross-lingual Cross-media Information Extraction and Event Tracking System.NAACL 2021.

Motivation

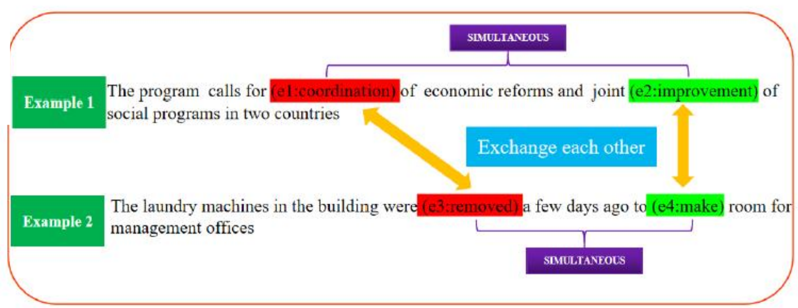
Label Imbalance Problem

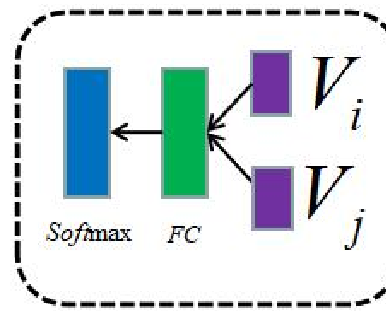
Limited amount of high-quality training data problem

$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t).$$

Weighting factor

Coefficient



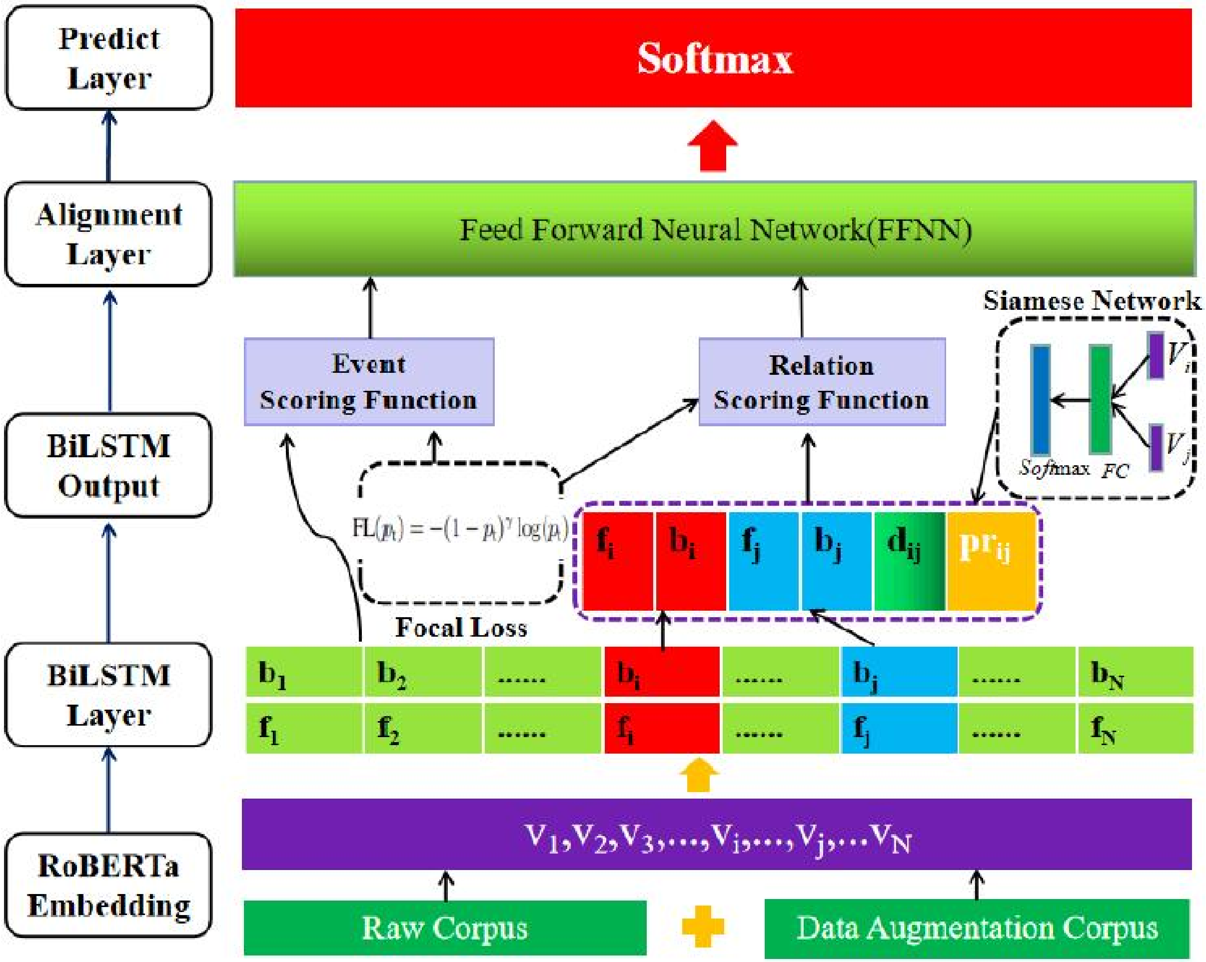


Focal Loss

Data Augmentation

Siamese Network

Our model



$$\mathcal{J}(\theta_1,\theta_2)=\alpha\sum_{i=1}^{|\mathcal{E}|}FL(e_i,\theta_1)+\beta\sum_{j=1}^{|\mathcal{EE}|}FL(r_j,\theta_2)+\gamma||\Phi||^2$$

Temporal Datasets

TB-Dense:
It is one of the basic data sets for temporal relation extraction, and contains three parts: TimeBank, AQUAINT and Platinum.

MATRES:
Ning Qiang et al proposed in 2018 that this dataset evolved and improved from TempEval3 dataset and had high IAA rate.

	TB-Dense	MATRES
# of Documents		
Train	22	183
Dev	5	-
Test	9	20
# of Pairs		
Train	4032	6332
Dev	629	-
Test	1427	827

Experiments

- Overall performance

Model(F1%)	TB-Dense		MATRES	
	Event	Relation	Event	Relation
CAEVO[3]	87.4	57.0	-	-
CogCompTime[12]	-	-	85.2	65.9
Perceptron[13]	-	-	-	69.0
RNN+CSE+ILP[2]	-	-	-	76.3
BiLSTM+MAP[4]	**89.2**	64.5	86.4	75.5
Our Model	88.1	**65.6**	**86.5**	**76.6**

Experiments

- Single relation performance

Model	BiLSTM+MAP			Our Model		
	P.	R.	F1.	P.	R.	F1.
B	70.4	58	63.6	**79.9**	58	**67.2**
A	61.0	69	64.7	**70.1**	66.7	**68.4**
I	23.8	9.0	13.0	**38.9**	12.5	**18.9**
II	40.0	30.1	34.4	46.7	26.4	33.7
S	-	-	-	-	-	-
V	61.2	67.2	64.1	61.4	**78.4**	**68.9**

BEFORE(B),AFTER(A),INCLUDES(I),IS INCLUDED(II),
SIMULTANEOUS(S),VAGUE(V).

Experiments

- Ablation Tests

Method(F1%)	TB-Dense	MATRES
All Components	65.6	76.6
w/o $Knowledge + Attention$	62.0	74.6
w/o $Attention$	62.3	75.1
w/o $DataAugmentation$	62.6	75.3
w/o $FocalLoss$	62.8	75.2

Conclusion and Future Work

- We propose a novel neural model for joint event and temporal relation extraction,which integrates temporal commonsense knowledge, data augmentation and FocalLoss function.

- The experimental results on two benchmark datasets show the effectiveness of our method and all three components.

- In the future, we will focus on solving the problem of few-shot labels.

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

Thank You!