Hierarchical Selection of Important Context for Generative Event Causality Identification with Optimal Transports

Hieu Man, Chien Van Nguyen, Linh Ngo and Thien Huu Nguyen

Presenter: Hieu Man

Outline

1. Introduction
2. Model
3. Experiments
4. Conclusion

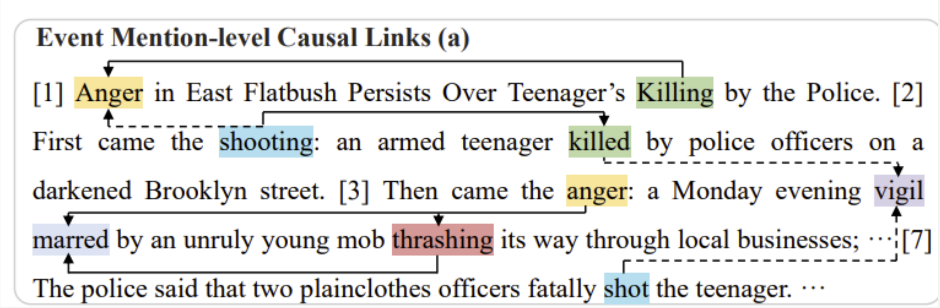
Introduction

Event Causality Identification (ECI)

- Event Causality Identification (ECI): Seek to predict causal relation between event mentions.

Document modeling in ECI

- Capture context between event pairs in the input document.
- Current methods use generation model like T5 to achieve state-of-the-art performance by auxiliary generating the important words.
- Issues:
  - Have limitations over the lengths.
  - The important context words for generation are obtained via dependency parsing tools or human annotation that are too noise or too expensive.



An example adopted from the EventStoryLine dataset
Chuang. et al 2022

Introduction

Our contributions

- Propose a novel generative model for ECI based on a hierarchical Optimal Transport alignment to select important contexts in the input document:
    - Formulate the identification of important contexts as an OT problem between event pairs and the document to automatically extract relevant contexts at both sentence and word levels
    - Learning paradigm to enhance important context extraction and ECI performance utilize RL and Generative model
    - Integrating external knowledge to further enhance the model.
- Conducting extensive experiments on different benchmark datasets over different languages to produce state-of-the-art performance for ECI

Model

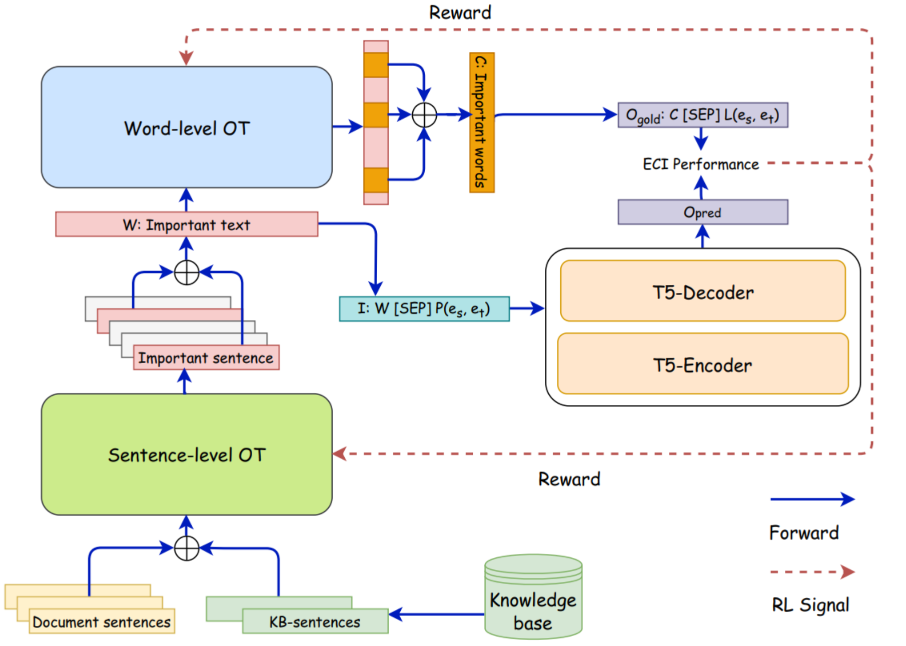


Figure 1: Illustration of our HOTECI framework for ECI. It consists of components for Sentence-level Optimal Transport (left-lower part), Word-level Optimal Transport (left-upper part), and Generative ECI Model (right part).

Model

- Sentence-level Optimal Transport
  - Solve the optimal transportation problem between two different set of sentences of document X and Y, the transportation cost T_{sent}

$$X = S \setminus \{s_s, s_t\} \ \ and \ \ Y = \{s_s, s_t, s_{null}\}$$

$$T_{sent}(x_i, y_j) = 1 - \text{cosine}(FF_1(\bar{x}_i), FF_1(\bar{y}_j))$$

  - With the transportation plan π_{sent}, the probability of the important score over the context sentences X is Q_{sent} :

$$\pi^*_{sent}(x_i) = \pi^*_{sent}(x_i, y_1) + \pi^*_{sent}(x_i, y_2)$$

$$Q_{sent}(x_i) = \text{softmax}(\pi^*_{sent}(x_i) | x_i \in X).$$

- Similarly, the probability of the important level over the context words U is:

$$Q_{word}(u_i) = \text{softmax}(\pi^*_{word}(u_i) | u_i \in U).$$

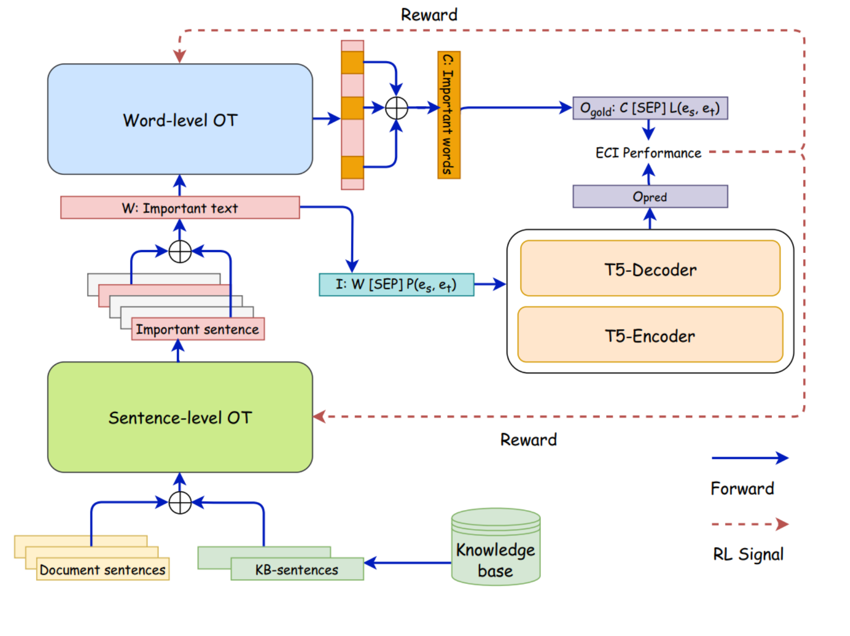


Figure 1: Illustration of our HOTECI framework for ECI. It consists of components for Sentence-level Optimal Transport (left-lower part), Word-level Optimal Transport (left-upper part), and Generative ECI Model (right part).

Model

- Generation objective
  - Reformulate the ECI problem into generation problem with W is the important context sentences and C is the important context words sampled from Q_{sent}, Q_{word}:

$$Input = W: Is\ there\ a\ causal\ relation\ between\ e_s\ and\ e_t$$

$$Output = C: Label(e_s, e_t)$$

  - Optimize the NLL loss:

$$L_{gen} = -\log P(O|I)$$

- Reinforce loss over the selection space with reward score R is the ECI performance:

$$\mathcal{L}_{RL} = -R \log P(W, C|S)$$

$$P(W, C|S) = \prod_{i=1}^{N_s} P(w_i|S) \prod_{j=1}^{N_w} P(c_j|W) = \prod_{i=1}^{N_s} Q_{sent}(w_i) \prod_{j=1}^{N_w} Q_{word}(c_j)$$

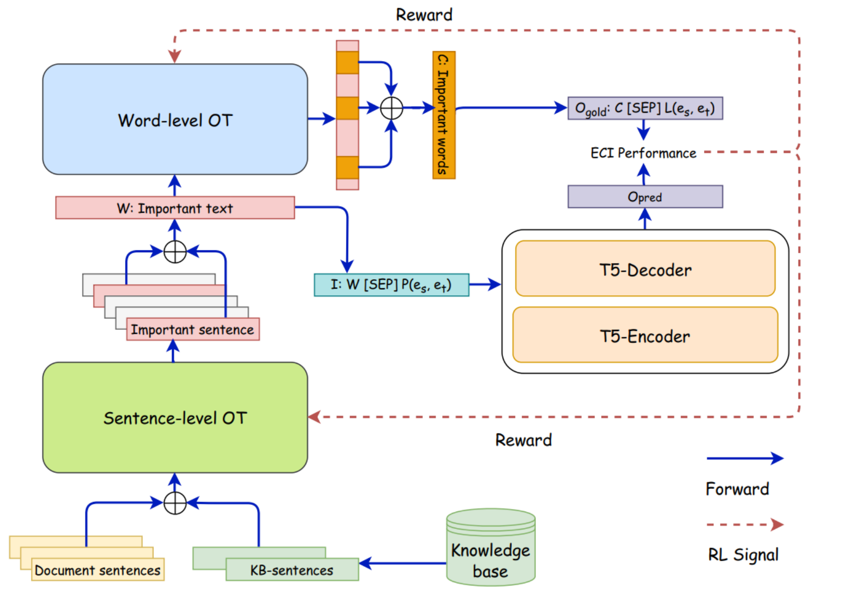


Figure 1: Illustration of our HOTECI framework for ECI. It consists of components for Sentence-level Optimal Transport (left-lower part), Word-level Optimal Transport (left-upper part), and Generative ECI Model (right part).

Experiments

Model	ESL (Intra-sentence)			ESL (Inter-sentence)			ESL (Intra + Inter)			CTB (Intra-Sentence)		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
LSTM (Gao et al., 2019)	34.0	41.5	37.4	13.5	30.3	18.7	17.6	33.9	23.2	-	-	-
Seq (Gao et al., 2019)	32.7	44.9	37.8	11.3	29.5	16.4	15.5	34.3	21.4	-	-	-
LR+ (Gao et al., 2019)	37.0	45.2	40.7	25.2	48.1	33.1	27.9	47.2	35.1	-	-	-
LIP (Gao et al., 2019)	38.8	52.4	44.6	35.1	48.2	40.6	36.2	49.5	41.9	-	-	-
ML (Mirza, 2014)	-	-	-	-	-	-	-	-	-	67.3	22.6	33.9
BERT (Tran and Nguyen, 2021)	39.2	49.3	43.7	22.3	29.2	25.3	27.3	35.3	30.8	38.5	43.9	41.0
KnowDis (Zuo et al., 2020)	39.7	66.5	49.7	-	-	-	-	-	-	42.3	60.5	49.8
Know (Liu et al., 2020)	41.9	62.5	50.1	-	-	-	-	-	-	36.6	55.6	44.1
RichGCN (Tran and Nguyen, 2021)	49.2	63.0	55.2	39.2	45.7	42.2	42.6	51.3	46.6	39.7	56.5	46.7
LearnDA (Zuo et al., 2021b)	42.2	69.8	52.6	-	-	-	-	-	-	41.9	68.0	51.9
CauSeRL (Zuo et al., 2021a)	41.9	69.0	52.1	-	-	-	-	-	-	43.6	68.1	53.2
ERGO-BERT (Chen et al., 2022)	49.7	72.6	59.0	-	-	-	-	-	-	58.4	60.5	59.4
ERGO-Longformer (Chen et al., 2022)	57.5	72.0	63.9	-	-	-	-	-	-	62.1	61.3	61.7
CF-ECI (Mu and Li, 2023)	47.1	66.4	55.1	-	-	-	-	-	-	50.5	59.9	54.8
CHEER (Chen et al., 2023)	59.9	69.9	62.6	45.2	52.1	48.4	49.7	53.3	51.4	56.4	69.5	62.3
SemSIn (Hu et al., 2023)	64.2	65.7	64.9	-	-	-	-	-	-	52.3	65.8	58.3
SENDIR (Yuan et al., 2023)	65.8	66.7	66.2	33	90	48.3	37.8	82.8	51.9	65.2	57.7	61.2
GenECI* (Man et al., 2022b)	58.7	65.7	61.9	-	-	-	-	-	-	58.6	59.3	58.6
DPJL (Shen et al., 2022)	65.3	70.8	67.9	-	-	-	-	-	-	63.6	66.7	64.6
HOTECI (ours)*	66.1	72.3	**69.1**	81.4	40.6	**55.1**	63.1	51.2	**56.5**	71.1	65.9	**68.4**

Table 1: Model's performance on ESL and CTB. The performance improvement of HOTECI over the baselines is significant with $p < 0.01$. * designates models that use T5.

Multilingual Experiment

Model	English			Danish			Spanish			Turkish			Urdu		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
PLM	48.7	59.9	53.7	35.9	36.2	36.0	50.6	49.1	49.9	44.0	59.4	50.5	40.4	43.2	41.8
Know	39.3	42.6	40.9	31.4	11.4	16.7	39.9	28.4	33.2	36.5	46.7	41.0	41.1	22.2	28.9
RichGCN	50.6	68.0	58.1	31.9	50.0	38.9	50.7	55.0	52.8	50.5	64.6	56.7	37.7	56.0	45.1
HOTECI (ours)	66.6	67.1	**66.8**	50.5	63.7	**56.3**	60.7	60.7	**60.7**	72.5	76.6	**74.5**	59.1	71.0	**64.5**

Table 2: Model's performance on MECI for different languages. The baselines use the base version of the multilingual RoBERTa model, i.e., XLMR (Conneau et al., 2020). PLM is similar to the BERT baseline in Table 1, but replaces BERT with XLMR.

Ablation study

#	Model	P	R	F1
1	HOTECI (full)	63.1	51.2	56.5
2	- RL loss	61.2	49.8	54.9
3	- Background Knowledge	62.2	50.5	55.7
4	- WS	61.1	47.5	53.4
5	- WS (five most similar words)	61.1	43.9	51.1
6	- WS (ten most similar words)	60.3	43.1	50.3
7	- SS (hosting sentences)	60.8	48.7	54.1
8	- SS (max surrounding sentences)	55.8	48.9	52.1
9	- SS (max most similar sentences)	61.2	50.1	55.1
10	- SS (five surrounding sentences)	58.9	47.6	52.7
11	- SS (five most similar sentences)	60.7	48.4	53.9
12	Selected Words to T5 Encoder	55.8	45.9	50.4
13	Uniform Dist for WS	63.2	49.8	55.7
14	Uniform Dist for SS	66.4	48.6	56.1

Table 3: Ablation study over test data of ESL using intra+inter sentence performance. WS and SS stand for word selection and sentence selection (respectively) with OT.

Conclusion

- ECI formulated as a generation problem
- Novel hierarchical method using optimal
- Handles document-level ECI with long contexts and cross-sentence events
- Achieves state-of-the-art performance on benchmarks
- Future work: Extend context selection for other NLP tasks

Thank you!