TAED: Topic-Aware Event Detection

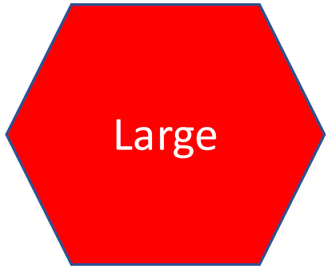
Yan Liang (yliang@ou.edu)

Christan Grant (cgrant@ou.edu)

https://oudatalab.com/

The UNIVERSITY of OKLAHOMA

Work Motivation



The amount of text generated every day is mind-blowing. Millions of data feeds are published and the ability to automatically organize and handle them is becoming indispensable.

Work Motivation

Across Domain

Large data leads to many diverse event types across different domains/topics (Ex: political events, fashion events ...)

Lack of techniques to handle event detection across domains.

Event Detection Applications

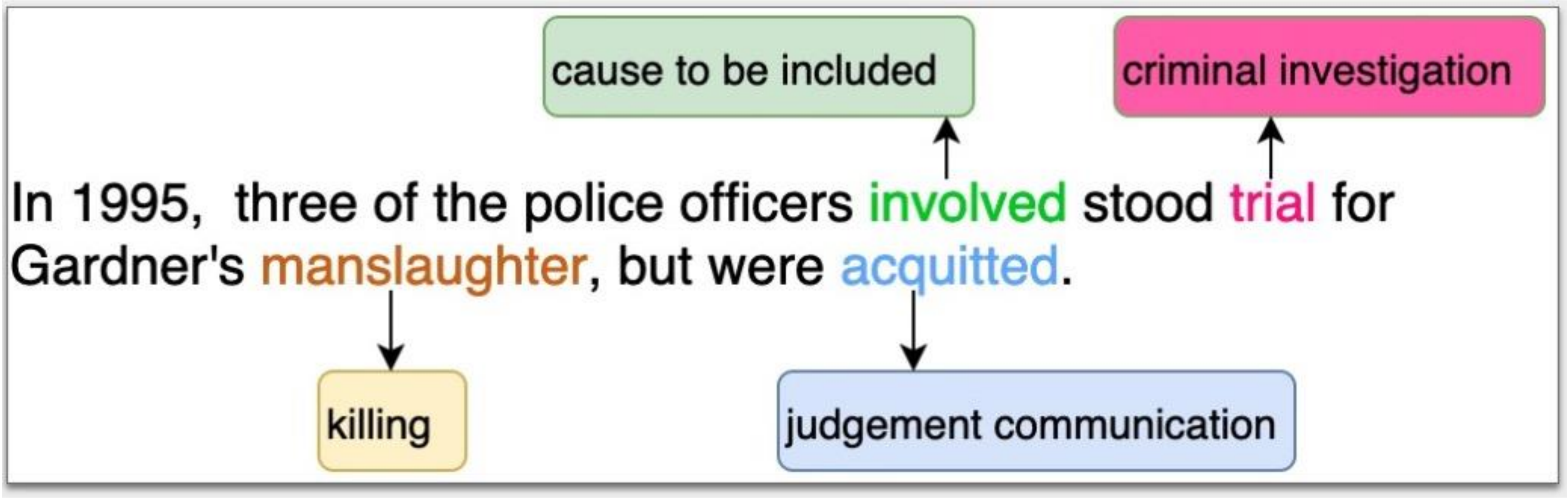
- Event Detection is one of the most important tasks in NLP domain.
  - Public affair management for government.
  - Help companies quickly discover market responses to their product.
  - Constructing or expanding the knowledge base.

Event Detection

Aims to find the event triggers --- the main word that most clearly expresses an event occurrence.

- Trigger Identification
- Trigger Classification

Event Detection



Challenges of Event Detection on More General Domains

- Lack of training data that covers variety of domains and event types.

- Training data is unbalanced on different event types. (variety of domains lead to variety of event types)

- Lack of methods to efficiently use domain/topic knowledge.

Motivation

- Semantically similar topics share similar event type distribution
- Semantically different topics have heterogeneous distributions of event types.

Different

Similar

earthquake	event types	catastrophe	causation	damaging	coming to be	destroying
	event types distribution	0.255	0.076	0.072	0.043	0.033
horse race	event types	competition	process start	process end	causation	hold
	event types distribution	0.201	0.104	0.058	0.047	0.036
terrorist attack	event types	attack	killing	terrorism	bodily harm	causation
	event types distribution	0.145	0.074	0.058	0.049	0.035
civilian attack	event types	killing	attack	statement	causation	bodily harm
	event types distribution	0.094	0.068	0.041	0.035	0.032

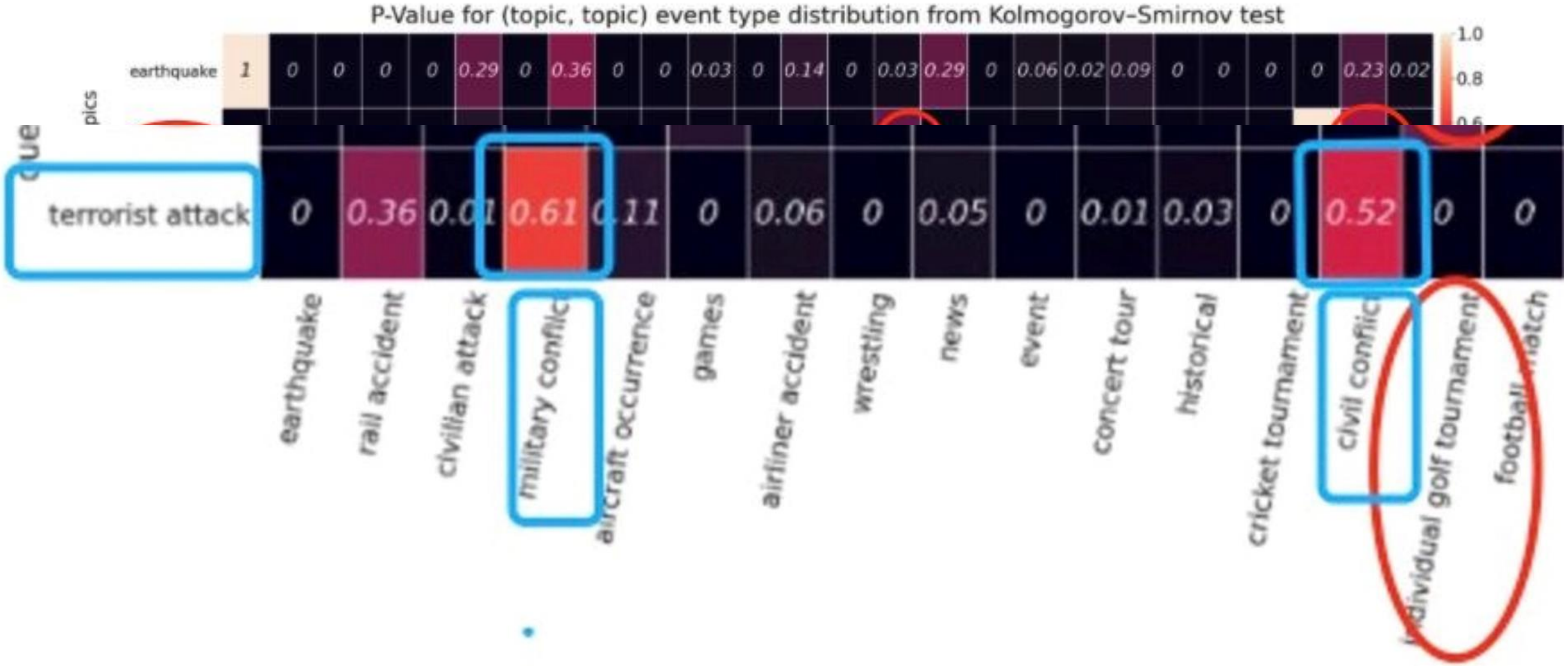
Table 4.1: An example of the top five event type distributions for each of the topics: earthquake, horse race, terrorist attack and civilian attack.

Motivation Explanation

- Each topic has 168 dimensional vectors to represent event type distributions.

- Using Kolmogorov-Smirnov test on two topics event type distribution ($P > 0.05$ means the null hypothesis that two topics follow the same event type can't be rejected)

Heuristic Explanation



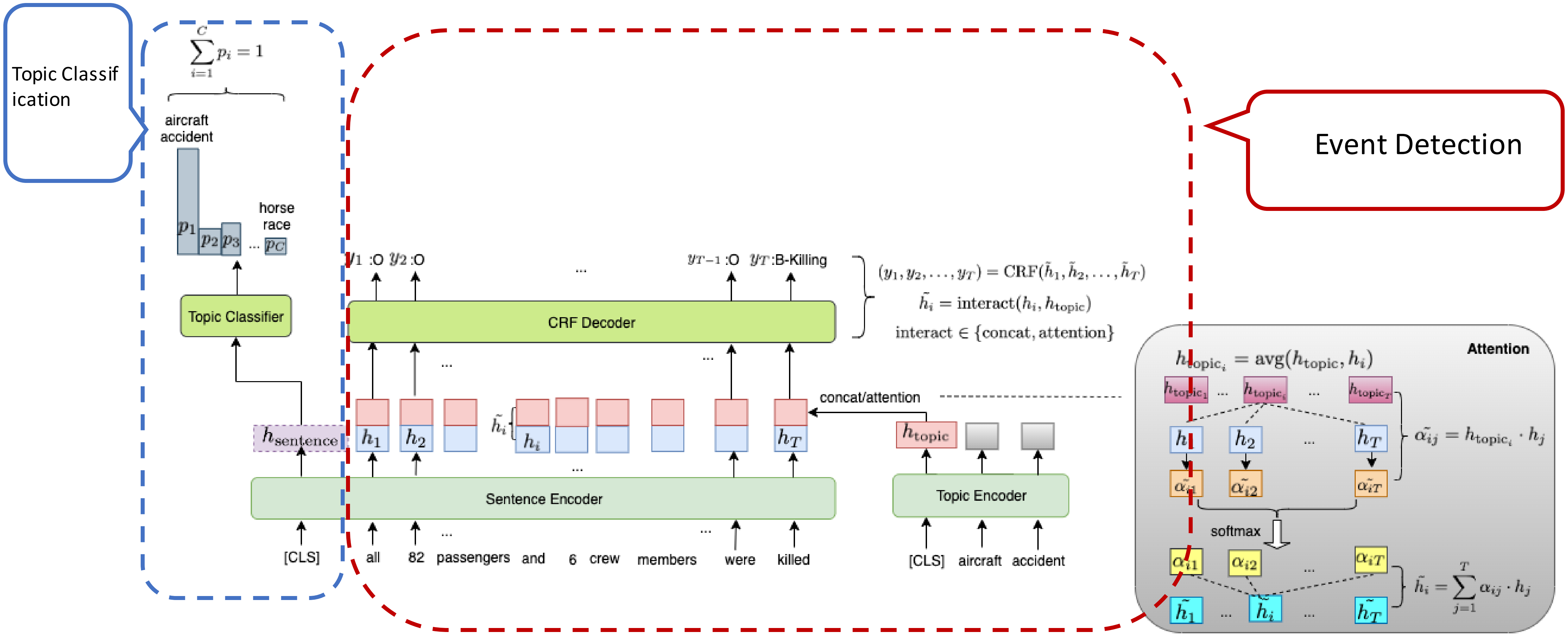
distributions between two topics. (Partial version. Full version in 4.3)

Horse race: International ice hockey competition (0.43), Individual golf tournament (0.29)
Terrorist attack: Military conflict (0.61), Civil conflict (0.52)

Our Contributions

- We perform detailed analysis explaining why topic helps on event detection task.
- We introduce topic name enhanced sentence representation for event detection & explore different ways to embed the topic name information.
- We train topic classification and event detection jointly.
- Topic-aware model achieves **+1.8%** F1 improvement on all event types and **+13.34%** F1 improvement on few-shot event types scenario compared to BERT baseline.

Model Architecture



Sentence and Topic Encoding

Sentence Encoding:

$$h_1, h_2 ..., h_T = \text{Encoder}(x_1, x_2, ..., x_T)$$

Topic Encoding: [CLS] token / average topic tokens embedding

$$h_{\text{topic}} = \text{TopicEncoder}(\text{topicword}_1, ..., \text{topicword}_N),$$

Topic-Aware Sentence Representation

- Topic-Aware Sentence Representation

  (1) concatenation:

$$\tilde{h} = (\tilde{h}_1, ..., \tilde{h}_T) = (h_1; h_{topic}, .., h_T; h_{topic}) \quad (4.3)$$

  (2) attention:

$$h_{topic_i} = average(h_{topic}, h_i)$$

$$\tilde{\alpha_{ij}} = h_{topic_i} \cdot h_j \quad (4.4)$$

$$(\alpha_{ij})_{j=1}^{T} = softmax((\tilde{\alpha_{ij}})_{j=1}^{T}) \quad (4.5)$$

$$\tilde{h_i} = \sum_{j=1}^{T} \alpha_{ij} \cdot h_j \quad (4.6)$$

Sequence Tagging Schema

BIOE tagging schema: Beginning, Inside, Outside, End.

Table 4.2: An example of the tag sequence for event type "Process Start" annotated with the BIOE scheme.

the	Total	Nonestop	Action	Wresting	(	TNA	)	promotion	that	took	place	on	October	23
O	O	O	O	O	O	O	O	O	O	B	E	O	O	O

Event Detection CRF Decoder

$$(y_1, y_2, ..., y_T) = CRF(\tilde{h}_1, \tilde{h}_2, ..., \tilde{h}_T) \quad (4.6)$$

The score for an input sequence X belongs to a specific topic to be assigned with a tag sequence Y can be calculated as:

$$score(X, topic, Y) = \sum_{i=1}^{T-1} \mathbf{T}_{y_i, y_{i+1}} + \sum_{i=1}^{T} \mathbf{E}_{i, y_i} \quad (4.7)$$

$$\mathbf{T} \in \mathbb{R}^{m \times m} \qquad \mathbf{E} \in \mathbb{R}^{T \times m}$$

Where m is the # of tags and T is the sequence length.

Event Detection Training

Log likelihood:

$$\log p(Y|X, topic) = \log \frac{score(X, topic, Y)}{\sum_{Y' \in tagset^T} score(X, topic, Y')} \quad (4.8)$$

Loss:

$$Loss_{event\_detection} = -\sum_{i=1}^{N} \log p(\hat{Y}_i | X_i, topic_i) \quad (4.9)$$

Topic Classification Training

$$(p_1, \ldots, p_C) = softmax(W_t \cdot h_{sentence} + b_t) \quad (4.10)$$

$$Loss_{topic} = -\sum_{j=1}^{N} \sum_{i=1}^{C} y_{ij} \log(p_{ij}) \quad (4.11)$$

C is the number of the topics in the training dataset

Multi-Task Training

$$Loss = Loss_{event\_detection} + \gamma \cdot Loss_{topic} \quad (4.11)$$

Where gamma is a non-negative tunable hyper-parameter

Performance (General)

Table 4.3: Performance on different datasets

Model Type	Dataset	P(%)	R(%)	F_1(%)
BERT-CRF	Full MAVEN Data	66.15 ± 0.24	69.64 ± 0.43	67.85 ± 0.07
BERT-CRF-TOPIC	Full MAVEN Data	66.28 ± 0.38	70.39 ± 0.40	$\mathbf{68.27 \pm 0.06}$
BERT-CRF	Generated Data 1	65.18 ± 1.14	70.32 ± 2.96	67.63 ± 0.05
BERT-CRF-TOPIC	Generated Data 1	66.21 ± 0.16	70.23 ± 0.16	$\mathbf{68.16 \pm 0.03}$
BERT-CRF	Generated Data 2	65.65 ± 0.30	69.74 ± 0.38	67.63 ± 0.08
BERT-CRF-TOPIC	Generated Data 2	66.35 ± 0.12	70.14 ± 0.34	$\mathbf{68.19 \pm 0.10}$
BERT-CRF	Generated Data topic-balanced 1	64.09 ± 1.67	62.68 ± 1.72	63.33 ± 0.05
BERT-CRF-TOPIC	Generated Data topic-balanced 1	63.9 ± 0.3	65.17 ± 0.18	$\mathbf{64.52 \pm 0.09}$
BERT-CRF	Generated Data topic-balanced 2	63.93 ± 1.51	63.21 ± 1.28	63.53 ± 0.11
BERT-CRF-TOPIC	Generated Data topic-balanced 2	64.41 ± 0.28	66.50 ± 0.24	$\mathbf{65.44 \pm 0.02}$
BERT-CRF	RAMS	34.05 ± 0.14	33.83 ± 0.03	33.94 ± 0.05
BERT-CRF-TOPIC	RAMS	36.67 ± 0.12	32.69 ± 0.04	$\mathbf{34.56 \pm 0.04}$

Better performance among all kinds of dataset

Event Type Groups

Groups	Occurrence	Event Type Count	Event Type Examples
Rare	(0,20]	38	besieging, ratification
Low	(20, 50]	35	warning, rescuing
Medium	(50, 100]	35	assistance, escaping
Sub-high	(100, 500]	53	damaging, destroying
High	$(500,\infty)$	7	catastrophe, causation

Table 4.4: Event type groups based on its occurrence frequency in training data

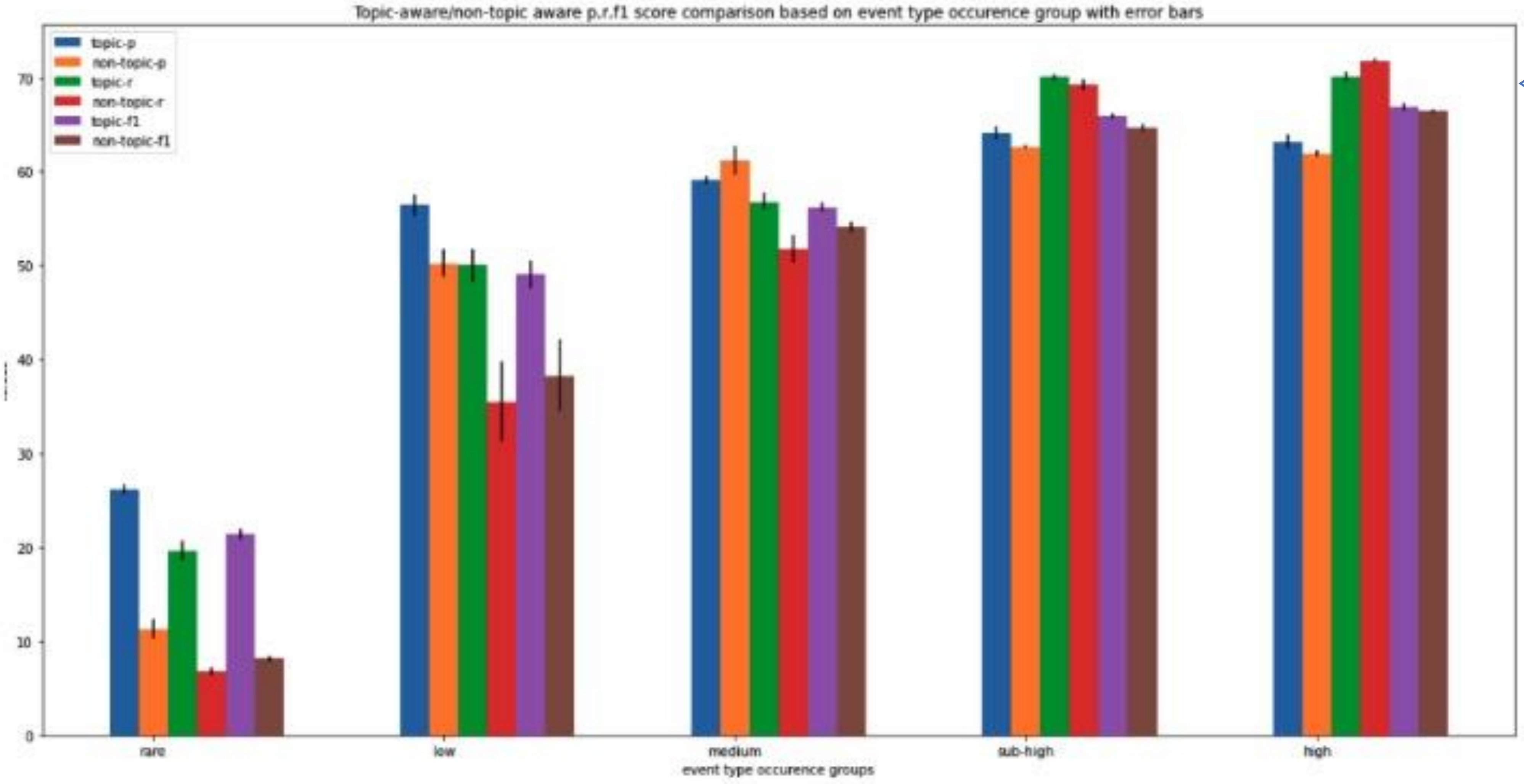


Figure 4.8: Topic-aware/non-topic-aware model Macro P, R, F1 performance with error bars on different event type occurrence groups defined in Table 4.4

Significant improvement "low resource" event types

Ablation Study 1: Topic Name Encoding

Table 4.7: Performance of using different ways to generate and utilize topic name embedding.

Model Type	**Topic Embedding Type**	**P(%)**	**R(%)**	**F_1(%)**
BERT-CRF-TOPIC	[CLS]	67	63.78	65.35
BERT-CRF-TOPIC	Average Token Embedding	64.53	66.44	65.47
BERT-CRF-TOPIC	[CLS] freeze	64.57	66.57	**65.56**
BERT-CRF-TOPIC	Average Token Embedding freeze	65	65.64	65.32
BERT-CRF-TOPIC	[CLS] (topic as attention)	63.93	67.02	65.44

No big performance diff among different ways to generate topic name embedding

Ablation Study 2: Topic Name Variations

1) Topic vocabulary (ranked by tf-idf feature) added

Table 4.8: Sample of 10 topic vocabulary terms and top-5 representative keywords.

Topic	Topic Vocabulary
earthquake	magnitude, occurred, quake, intensity, damage
winter storm	snow, blizzard, snowfall, new, winds
tennis event	open, doubles, slam, singles, djokovic
rugby match	chiefs, brumbies, sharks, final, crusaders
university boat race	oxford, cambridge, lengths, crews, goldie
war	paulo, vargas, 1930, presets, garais
military operation	bomb, manchester, ira, bombing, embassy
swimming event	golds, medals, bronze, silver, freestyle pool
cricket series	ashes, england, australia, test, wickets
civilian attack	massacre, attack, kill, police, people

2) General event name removed ---> make topic embedding more discriminative

- recurring event --> recurring
- historical event --> historical
- wrestling event --> wrestling

Ablation Study 2: Topic Name Variations

Table 4.6: TAED performance with different topic-classification weights, performance of general event words kept/removed and performance of extra topic keywords added on for a specific topic.

Model Type	topic-classification weight	general event word removed	P(%)	R(%)	F_1(%)
BERT-CRF	NA	NA	66.91	60.71	63.66
BERT-CRF-TOPIC	1	True	64.44	66.52	**65.46**
BERT-CRF-TOPIC	0	True	63.52	66.04	64.73
BERT-CRF-TOPIC	0.1	True	62.37	66.66	64.44
BERT-CRF-TOPIC	0.5	True	64.17	64.92	64.53
BERT-CRF-TOPIC	2	True	63.76	65.33	64.51
BERT-CRF-TOPIC	10	True	64.26	58.73	61.38
BERT-CRF-TOPIC	25	True	63.8	47.34	54.33
BERT-CRF-TOPIC	50	True	60.36	33.06	42.65
BERT-CRF-TOPIC	75	True	55.98	25.69	34.99
BERT-CRF-TOPIC	100	True	49.58	19.38	27.79
BERT-CRF-TOPIC	1	False	65.59	64.29	64.93
BERT-CRF-TOPIC (with vocab)	1	True	64.97	65.08	65.02

General event word removed

Topic vocabulary added

Ablation Study 3: Multi-Task Learning

Table 4.6: TAED performance with different topic-classification weights, performance of general event words kept/removed and performance of extra topic keywords added on for a specific topic.

Model Type	topic-classification weight	general event word removed	P(%)	R(%)	F_1(%)
BERT-CRF	NA	NA	66.91	60.71	63.66
BERT-CRF-TOPIC	1	True	64.44	66.52	**65.46**
BERT-CRF-TOPIC	0	True	63.52	66.04	64.73
BERT-CRF-TOPIC	0.1	True	62.37	66.66	64.44
BERT-CRF-TOPIC	0.5	True	64.17	64.92	64.53
BERT-CRF-TOPIC	2	True	63.76	65.33	64.51
BERT-CRF-TOPIC	10	True	64.26	58.73	61.38
BERT-CRF-TOPIC	25	True	63.8	47.34	54.33
BERT-CRF-TOPIC	50	True	60.36	33.06	42.65
BERT-CRF-TOPIC	75	True	55.98	25.69	34.99
BERT-CRF-TOPIC	100	True	49.58	19.38	27.79
BERT-CRF-TOPIC	1	False	65.59	64.29	64.93
BERT-CRF-TOPIC (with vocab)	1	True	64.97	65.08	65.02

Multi-task learning is helpful

Figure 4.9: F_1 performance vs. γ (Each γ on X-axis has been run 5 times with different random seeds represented by points with different colors. The curve is the average performance of the 5 runs for each γ.)

Case Study

Topic-Aware can help on both Trigger identification and Trigger Classification.

- The trigger words identified is wrong
- Though the trigger words are correct, but the event type is wrong

Table 4.9: Performance of BERT-CRF and BERT-CRF-TOPIC only on Trigger Identification

Model Type	**P(%)**	**R(%)**	$\mathbf{F_1(\%)}$
BERT-CRF	77.3	77.9	77.6
BERT-CRF-TOPIC	77.93	78.59	**78.26**

Case Study

Flight 821 is the deadliest accident involving a Boeing 737-500, surpassing the 1993 crash of Asiana Airlines Flight 733, and was the second-deadliest aviation incident in 2008, behind Spanair Flight 5022.

Topic: aircraft accident

Top topic event types: catastrophe, causation, motions.

Non-topic-aware predictions: None

Topic-aware predictions: accident (catastrophe), incident (catastrophe)

Case Study

This was the first southern stadium rock show since ZZ TOP played to 80,000 people at UT Austin on September 1, 1974 and tore up the field.

Topic: music festival

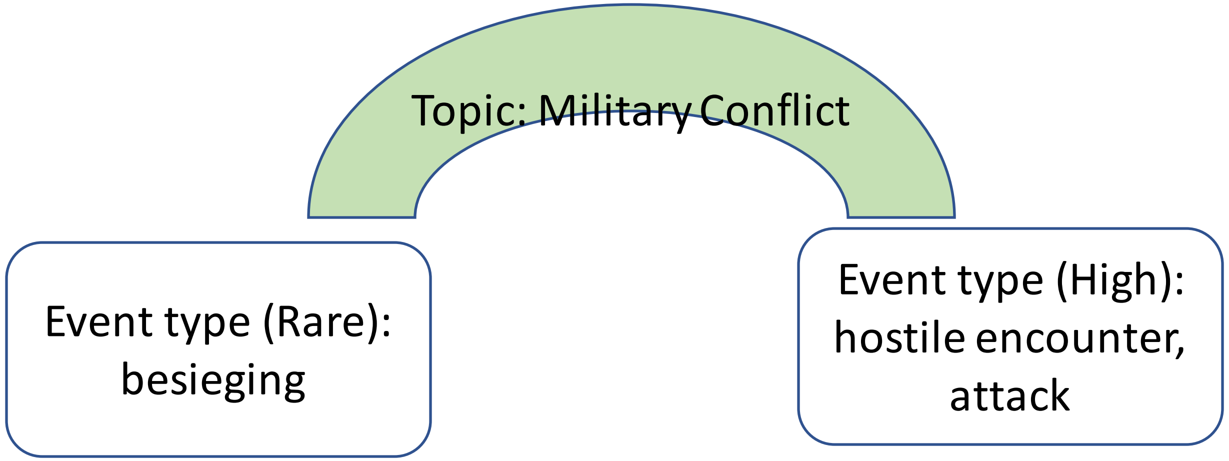
Top topic event types: social event, process start, arranging, competition.

Non-topic-aware predictions: played (B-participation)

Topic-aware predictions: played (B-competition)

Case Study

- How topics work as a bridge to connect "low resource" and "high resource" event types?



Thank you!

The UNIVERSITY of OKLAHOMA