# TacoERE: Cluster-aware Compression for Event Relation Extraction

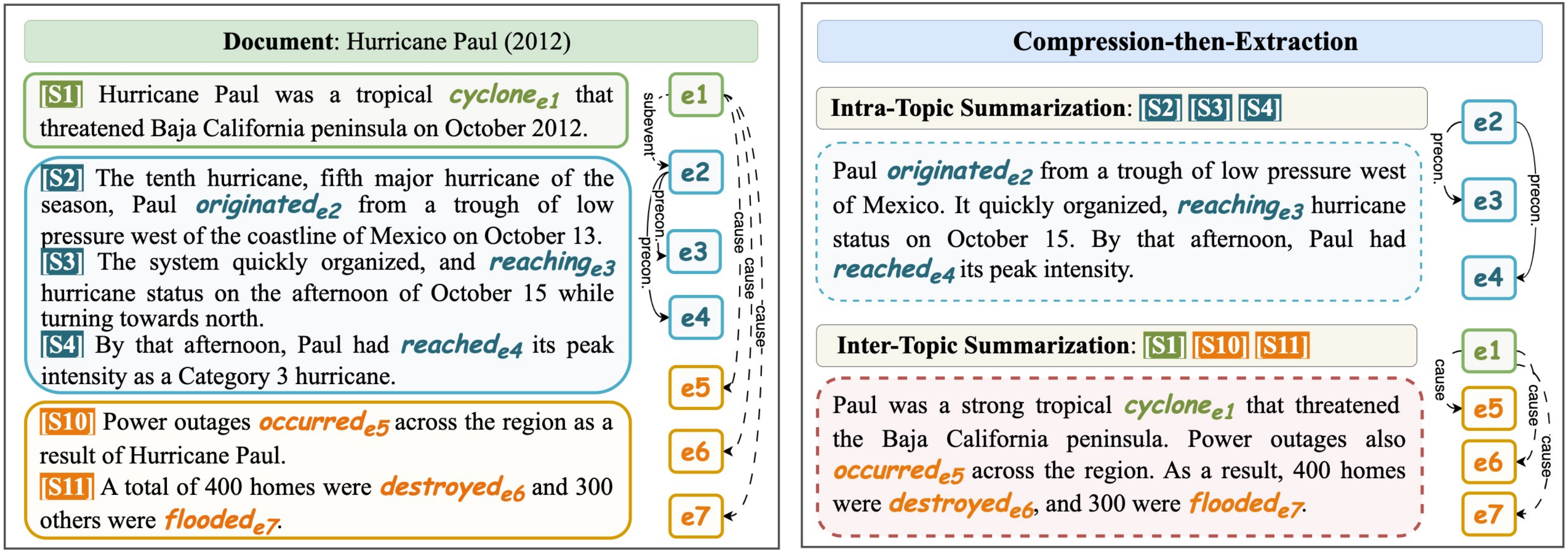
Yong Guan, Xiaozhi Wang, Lei Hou, Juanzi Li, Jeff Z. Pan,
Jiaoyan Chen, Freddy Lecue

# Outline

- Introduction
- Approach
- Experiments
- Conclusions

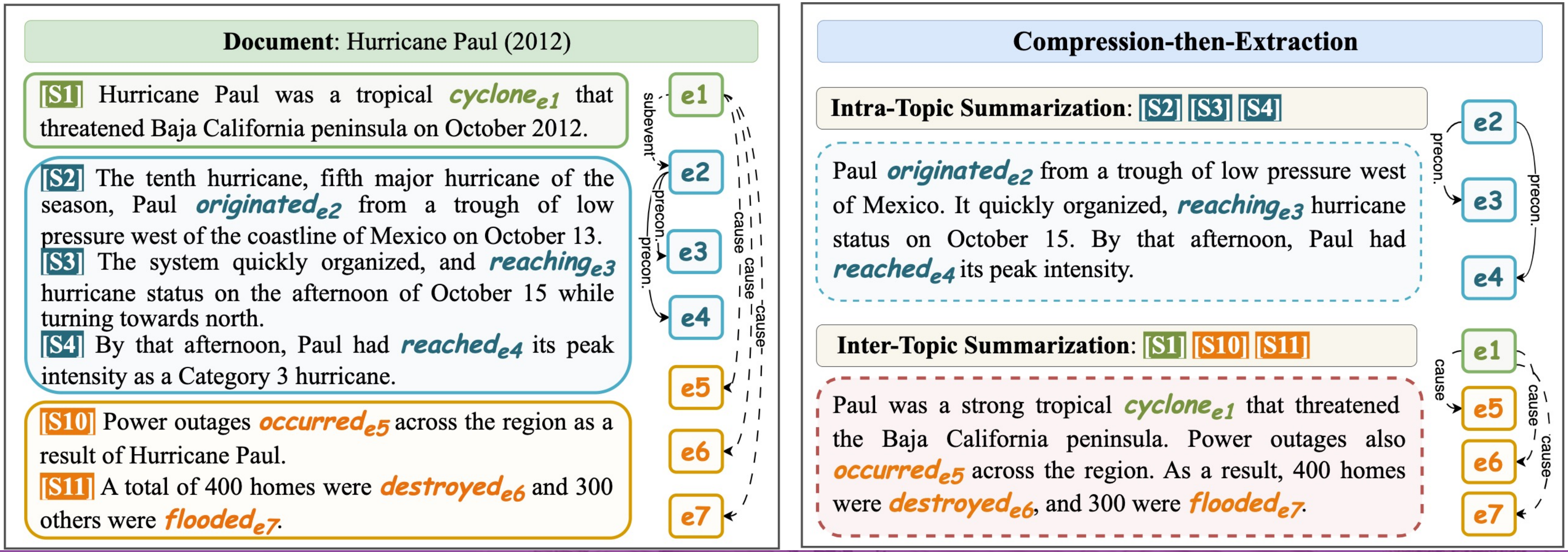
# Introduction—Definition

- **Event Relation Extraction (ERE)** aims to predict relations, such as causal and subevent relations, between event mentions or trigger words in a document



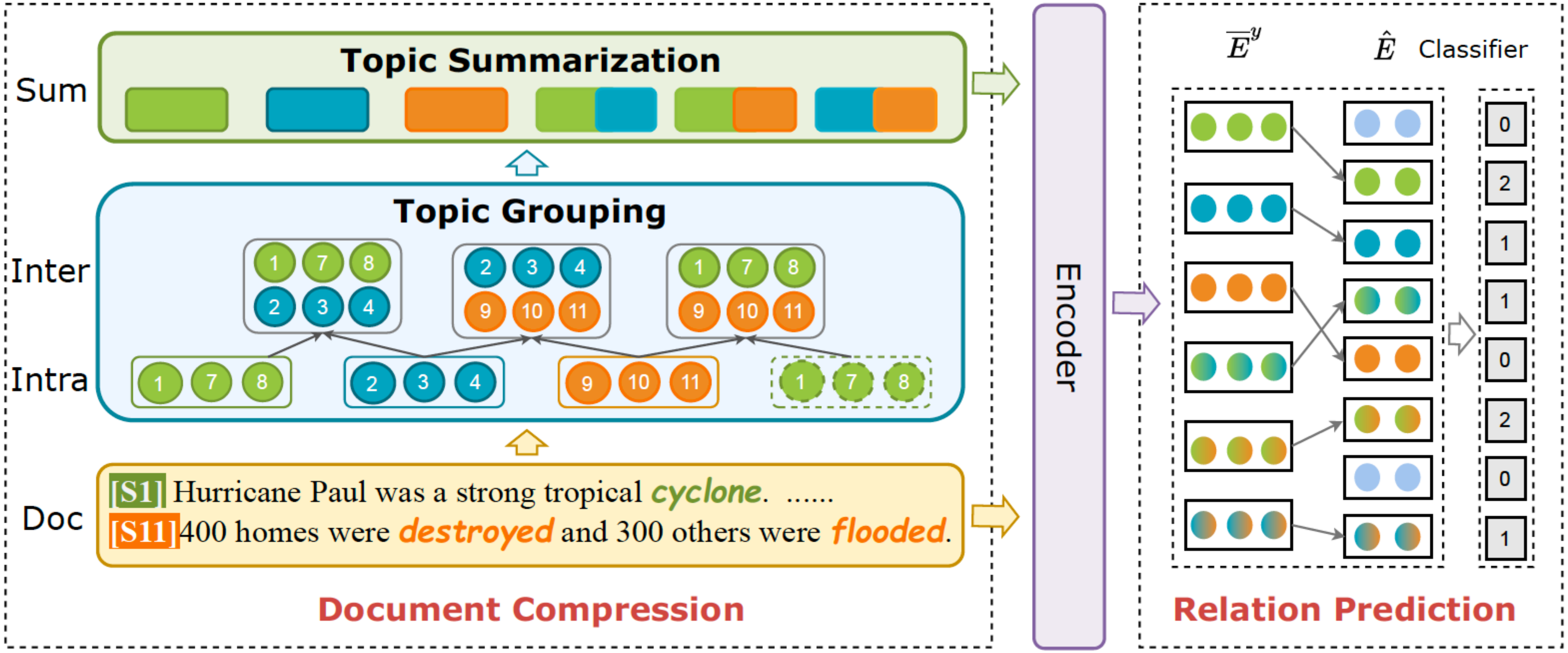
# Introduction—Challenges

- **Long-Range Dependencies** indicates events may be scattered across multiple sentences
- **Information Redundancy** refers to the existence of information non relevant for relation prediction



# Approach

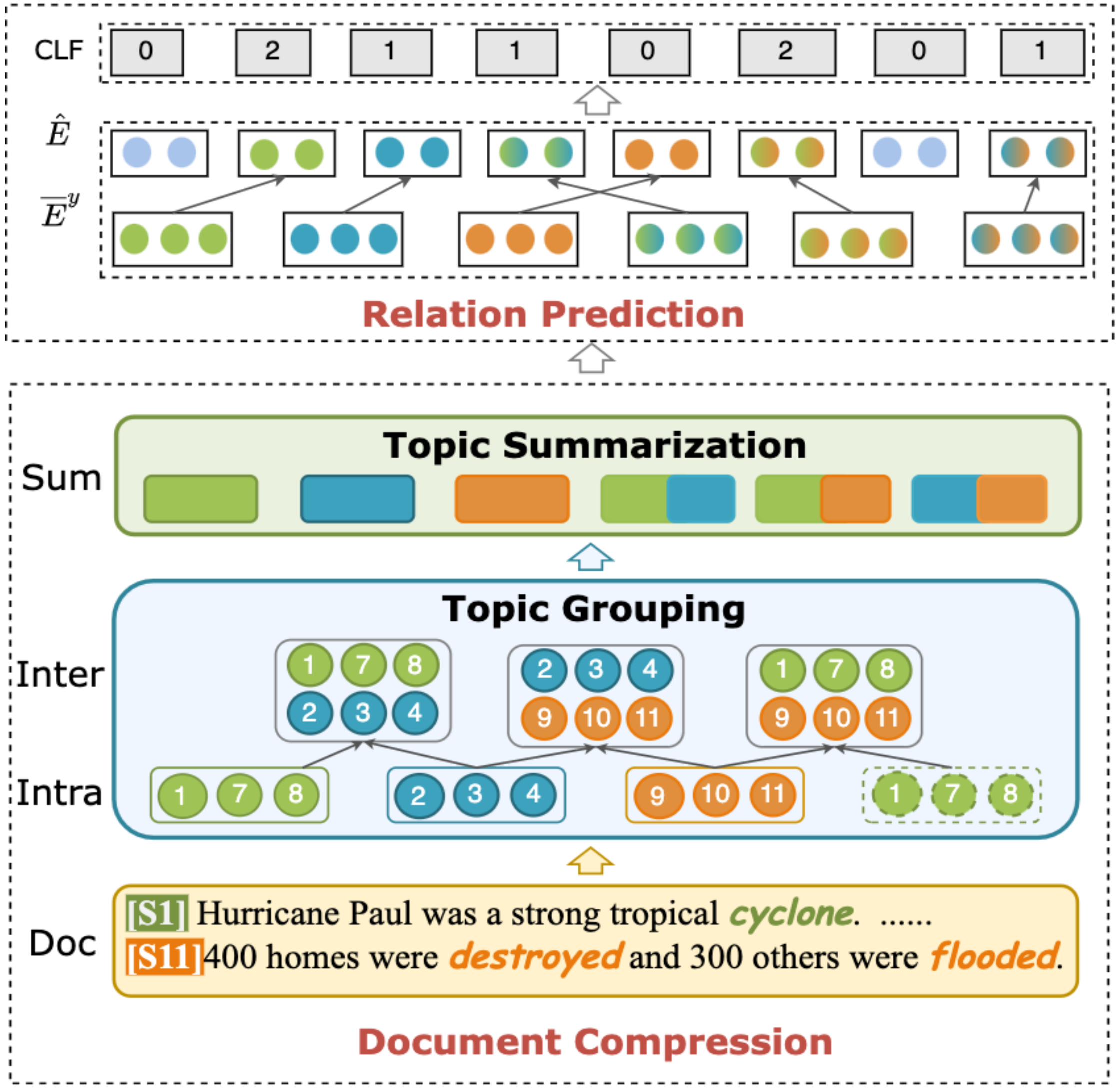
- **TacoERE**: A cluster aware compression method for improving ERE, which explores a compression then-extraction paradigm to extract event relations



# Approach

- **TacoERE**: Implementation on small scale pre-trained language models (PLMs), such as RoBERTa

- Document Clustering
- Cluster Summarization
- Relation Prediction
- Joint Training

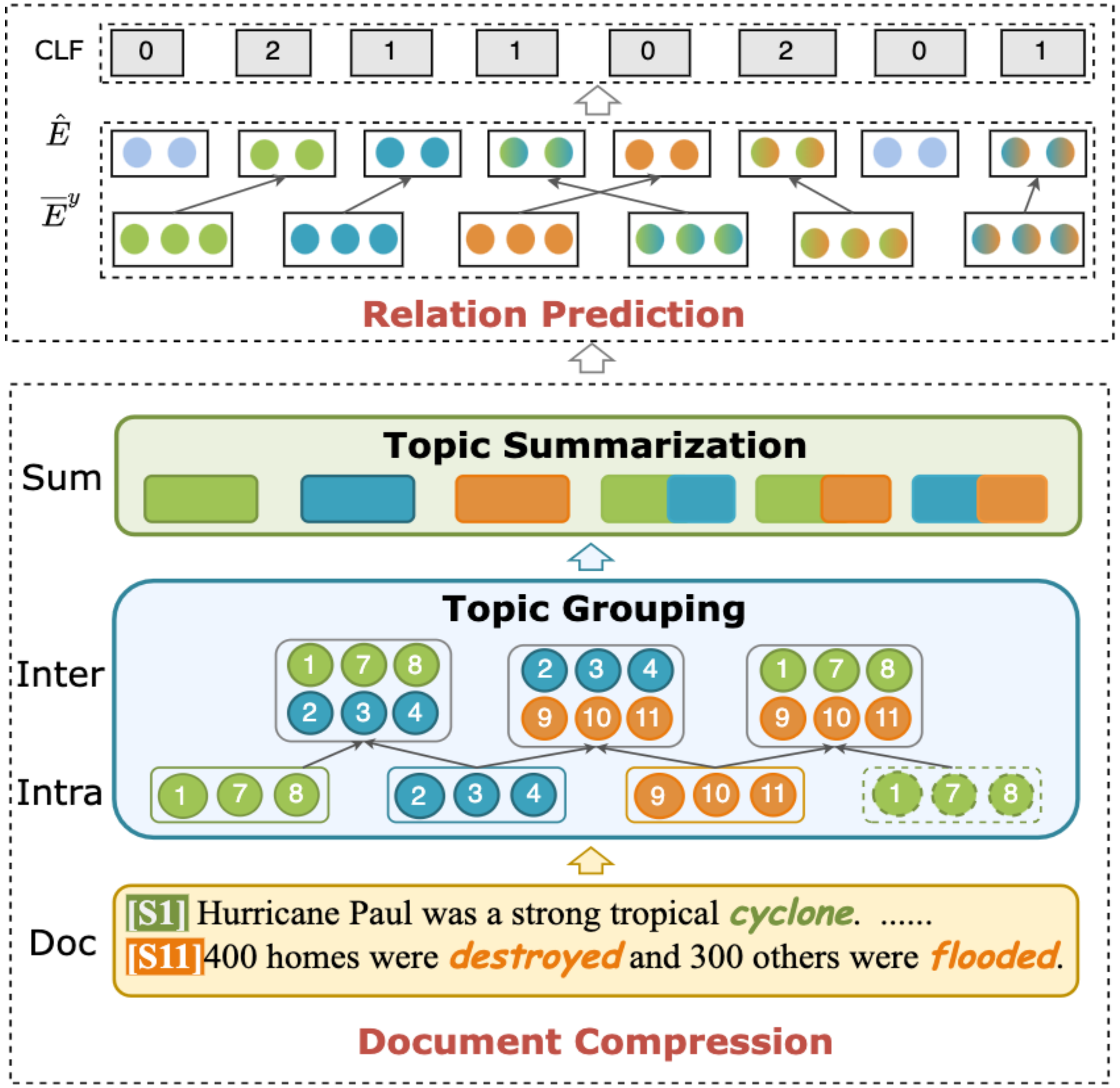


# Approach

- **Document Clustering** aims to reduce the distance between events by splitting document into intra- and inter-clusters
  - **Intra-Clusters** aim to enhance the relations within the same cluster

$$J=\sum_{i=1}^{k}\sum_{j=1} p_{ij}||v_j-u_i||^2$$

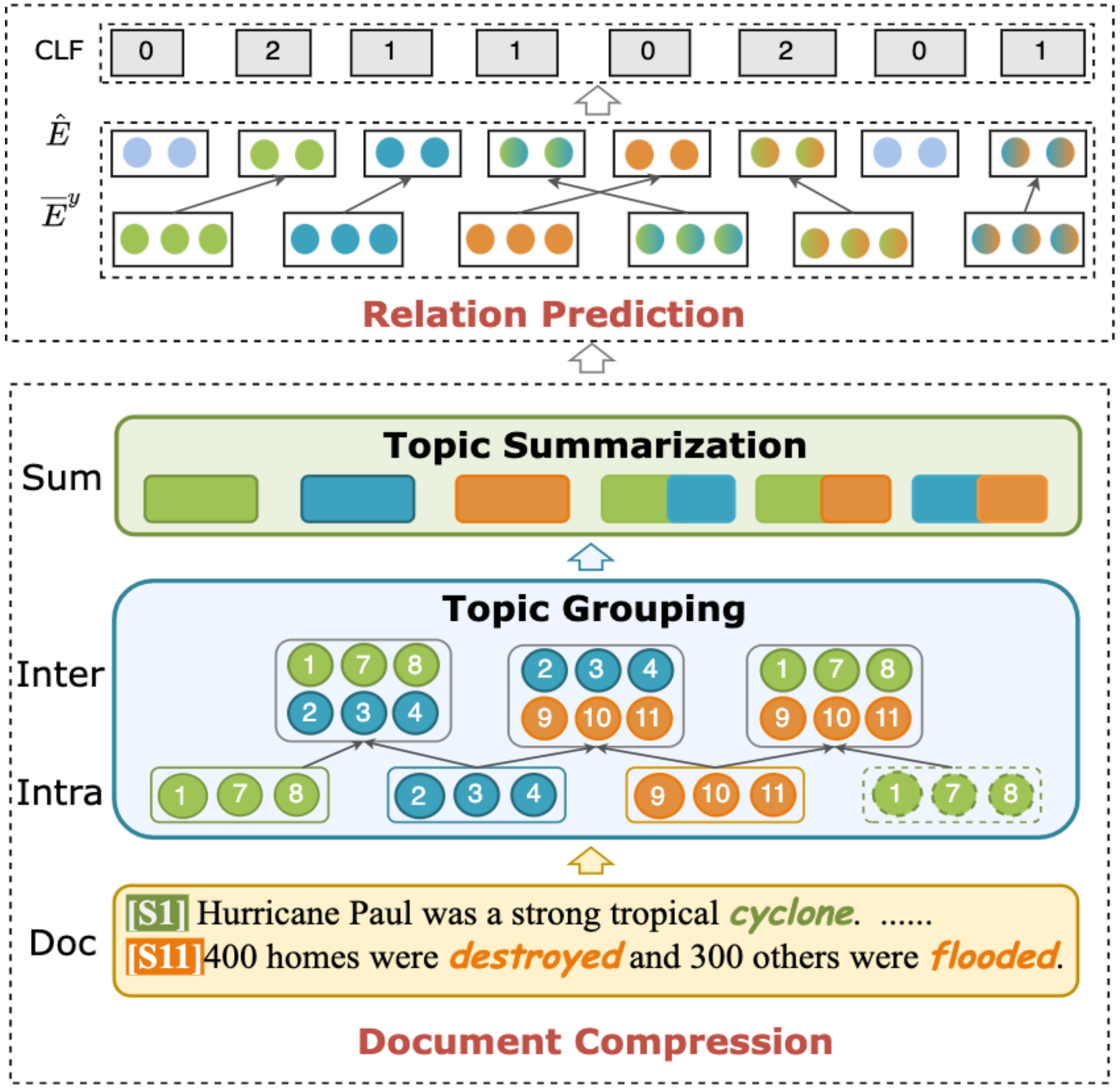
  - **Inter-Clusters** attempt to model the related events at arbitrary distances.



# Approach

- **Cluster Summarization** aims to simplify and highlight important text content of clusters
- **Relation Prediction** aims to predict the relations based on the text content from cluster summarization

$$\mathcal{L}_{rp} = -\sum_{i \neq j} \sum_{\mathcal{R}} \{ r_{ij} log P_{ij} + (1 - r_{ij}) log(1 - P_{ij}) \}$$



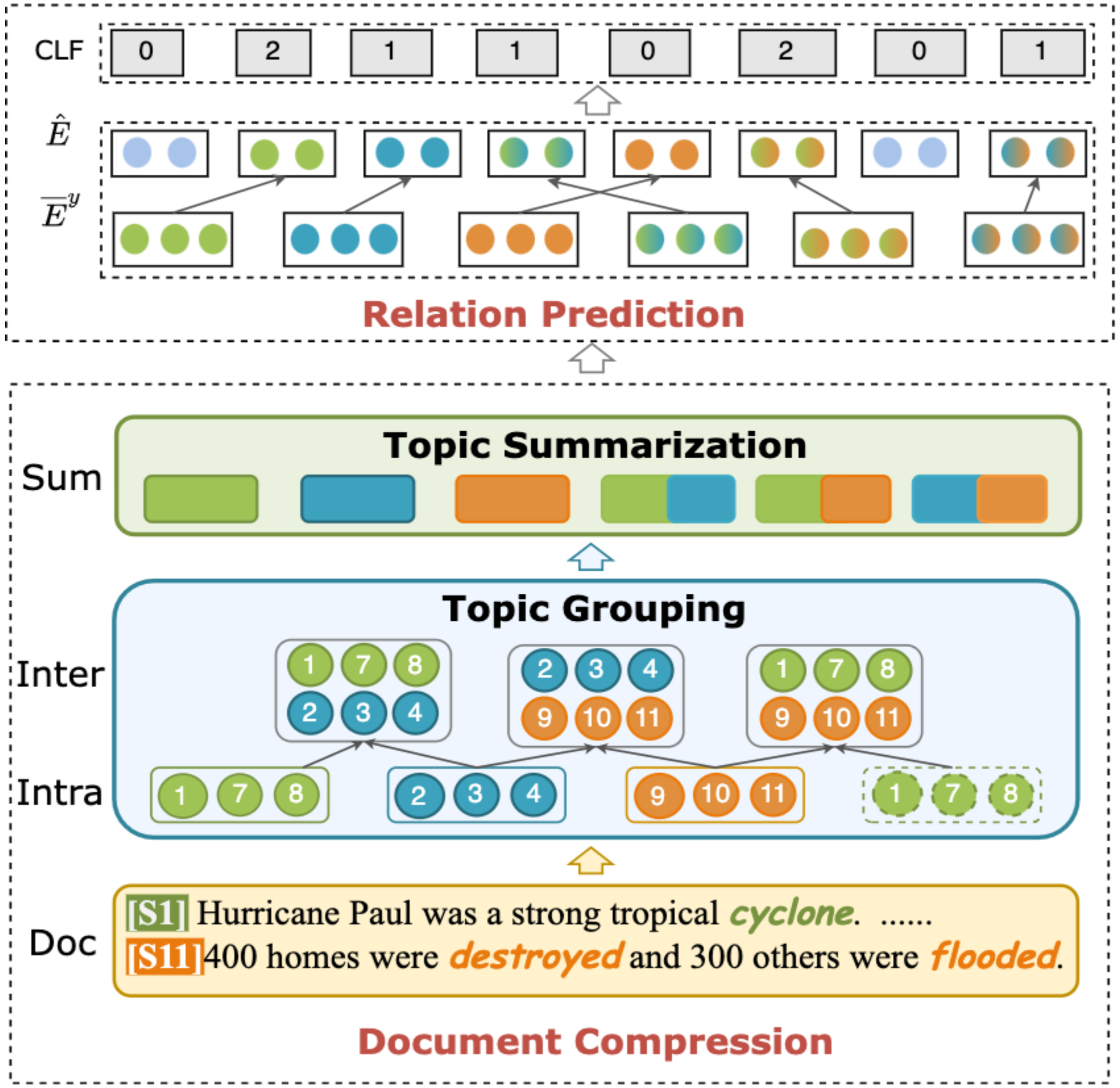
# Approach

- **Joint Training** aims to jointly optimize the cluster summarization and relation prediction

$$\mathcal{R}(C) = \alpha \mathcal{R}^{per}(C) + \beta \mathcal{R}^{ec}(C)$$

$$\mathcal{L}_{sum} = -\mathbb{E}_{C' \sim P(C'|e_i, e_j, D)}[\mathcal{R}(C')]$$

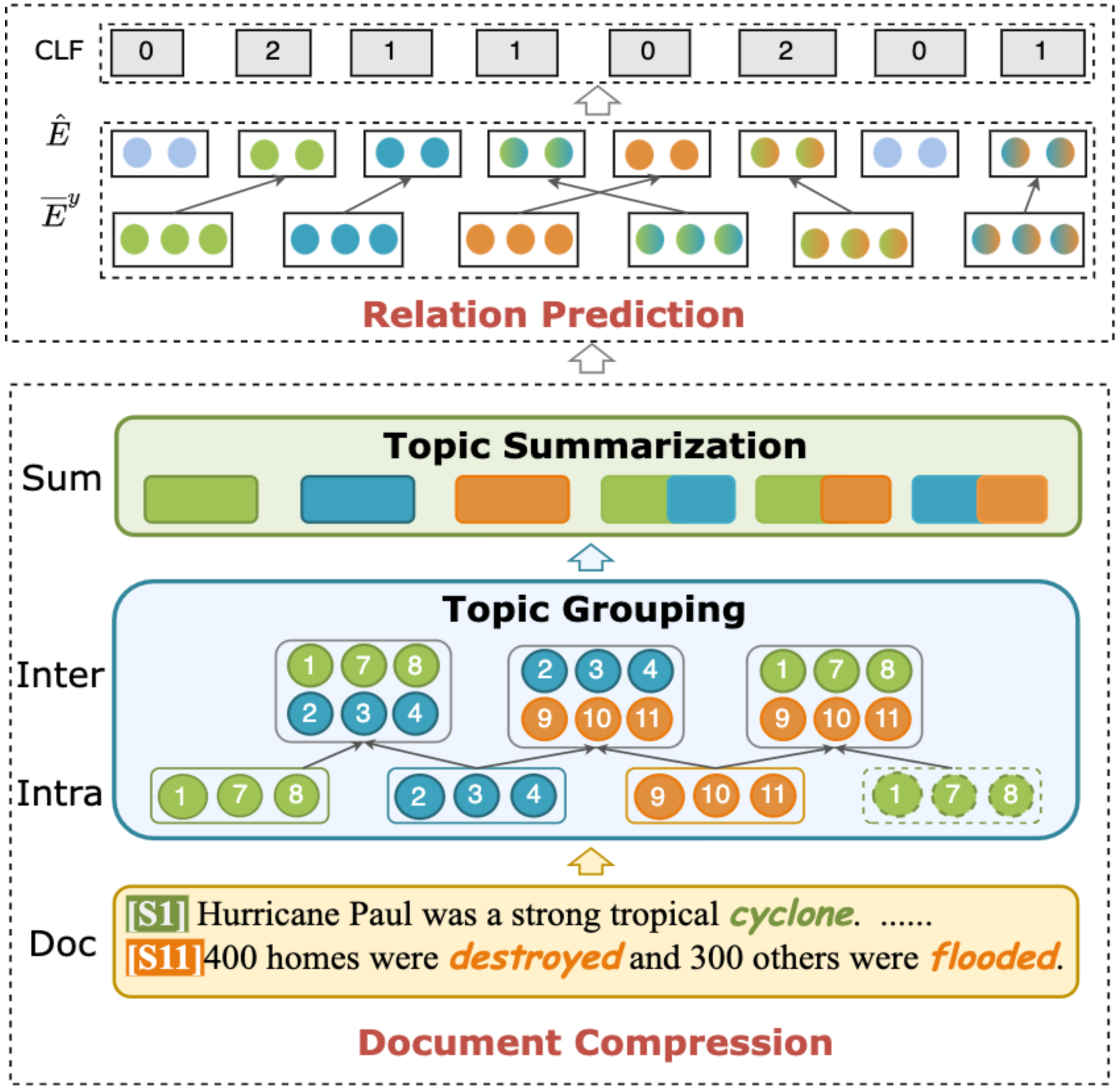
$$\nabla \mathcal{L}_{sum} = -(\mathcal{R}(C) - \theta) \nabla log P(C|e_i, e_j, D)$$



# Approach

- **TacoERE**: Implementation on large language models (LLMs), such as ChatGPT and GPT-4

- Document Clustering
- Cluster Summarization
- Relation Prediction



# Experiments

## Datasets

- MAVEN-ERE
- HiEve
- EventStoryLine

| Datasets | Documents | Evens | Temporal | Causal | Subevent |
|---|---|---|---|---|---|
| MAVEN-ERE | 4,480 | 112,276 | 1,216,217 | 57,992 | 15,841 |
| HiEve | 100 | 2,734 | - | - | 3,648 |
| EventStoryLine | 258 | 4,732 | 8,111 | 4,584 | - |

# Experiments

- Performance on small scale pre-trained language models

| Methods | P | R | F1 |
|---|---|---|---|
| MAVEN-ERE | | | |
| BERT | 31.6 | 28.2 | 29.9 |
| RoBERTa | 33.8 | 29.5 | 31.5 |
| Hierarchical | 31.8 | 29.2 | 30.6 |
| SIEF | 33.6 | 30.8 | 32.3 |
| TacoERE (PLMs) | **34.8** | **32.4** | **34.1** |
| EventStoryLine | | | |
| BERT | 30.3 | 9.4 | 12.8 |
| RoBERTa | 31.1 | 10.7 | 14.4 |
| Hierarchical | 30.1 | 10.2 | 13.1 |
| SIEF | 32.4 | 11.3 | 14.8 |
| SCS-EERE | 32.7 | 10.9 | 15.1 |
| TacoERE (PLMs) | **32.9** | **12.3** | **16.4** |

Results of Causal Relations

| Methods | P | R | F1 |
|---|---|---|---|
| MAVEN-ERE | | | |
| BERT | 27.5 | 24.7 | 26.8 |
| RoBERTa | 29.8 | 25.6 | 27.5 |
| Hierarchical | 28.4 | 25.4 | 27.1 |
| SIEF | 30.2 | 26.4 | 28.7 |
| TacoERE (PLMs) | **31.8** | **28.9** | **30.6** |
| HiEve | | | |
| BERT | 19.8 | 15.2 | 16.3 |
| RoBERTa | 20.2 | 16.1 | 17.8 |
| Hierarchical | 21.4 | 17.3 | 16.7 |
| SIEF | 21.8 | 17.4 | 18.6 |
| SCS-EERE | 20.6 | **19.7** | 19.2 |
| TacoERE (PLMs) | **22.6** | 19.5 | **20.8** |

Results of Sub-Event Relations

# Experiments

- Performance on large language models (LLMs)

| Methods | Text-Davinci-003 | | | ChatGPT | | | GPT-4 | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 | P | R | F1 |
| Document | 13.8 | 6.2 | 8.5 | 21.7 | 32.2 | 25.9 | 27.1 | 41.5 | 32.8 |
| Sentence Pair | 21.9 | 7.1 | 10.7 | 24.3 | 31.2 | 27.3 | 33.4 | 38.6 | 35.7 |
| Topic Grouping | 17.3 | 8.1 | 10.9 | 24.6 | 32.9 | 28.2 | 31.9 | 47.1 | 38.1 |
| TacoERE (LLMs) | **30.2** | **8.9** | **13.8** | **31.3** | **45.6** | **37.1** | **38.9** | **45.5** | **41.9** |

Table 5: Model performance of causal relation on different LLMs. Results are obtained under 2-shot setting.

# Experiments

## Case Study

**Document:** [S1] The Harrow and Wealdstone rail crash was a three-train <Event> **collision** </Event> at Harrow and Wealdstone station in Wealdstone , Middlesex during the morning rush hour of 8 October 1952 ; 112 were killed and 340 injured; it remains the worst peacetime rail crash in the UK.
[S2] An overnight express train from Perth crashed at speed into the rear of a local passenger train standing at a platform at the station .
[S3] The wreckage blocked adjacent lines and was struck within seconds by a `` double-headed " express train travelling north at .
[S4] The Ministry of Transport report on the crash found that the driver of the Perth train had passed a caution signal and two danger signals before colliding with the local train .
[S6] The accident <Event> **accelerated** </Event> the introduction of Automatic Warning System – by the time the report on the accident had been published , British Railways had agreed to a five-year plan to install the system to give drivers an in-cab audible and visual warning of being about to pass a distant signal at caution.

Text-Davinci-003: ✗ ChatGPT: ✗ GPT-4: ✗

**Ground Truth:** PRECONDITION

**Sentence Pair:** [S1], [S6] — Text-Davinci-003: ✗ ChatGPT: ✗ GPT-4: ✗

**Topic Grouping:** [S1], [S2], [S3], [S6] — Text-Davinci-003: ✗ ChatGPT: ✗ GPT-4: ✓

**TacoERE (LLMs):** The Harrow and Wealdstone rail crash, a devastating <Event> **collision** </Event> involving three trains, occurred at Harrow and Wealdstone station. The crash resulted in the loss of 112 lives and injuries to 340 individuals, making it the worst peacetime rail crash in the UK. The accident prompted the <Event> **accelerated** </Event> introduction of the Automatic Warning System, ensuring caution when approaching distant signals.

Text-Davinci-003: ✗ ChatGPT: ✓ GPT-4: ✓

# Conclusions

- We propose a novel cluster-aware compression method for event relation extraction, namely, TacoERE, which explores a compression-then-extraction paradigm to extract relations.
- We utilize document clustering to split the document into intra- and inter-clusters to allow the modeling of dependencies without any reliance on event distance.
- Experimental results show that our proposed TacoERE outperforms existing methods, especially on LLMs, with improvements by 11.2% and 9.1% on ChatGPT and GPT-4 respectively.


# Thanks for your listening !