Part IV: Language Models for Knowledge Base Construction

KDD 2023 Tutorial

Pretrained Language Representations for Text Understanding: A Weakly-Supervised Perspective

Yu Meng, Jiaxin Huang, Yu Zhang, Yunyi Zhang, Jiawei Han

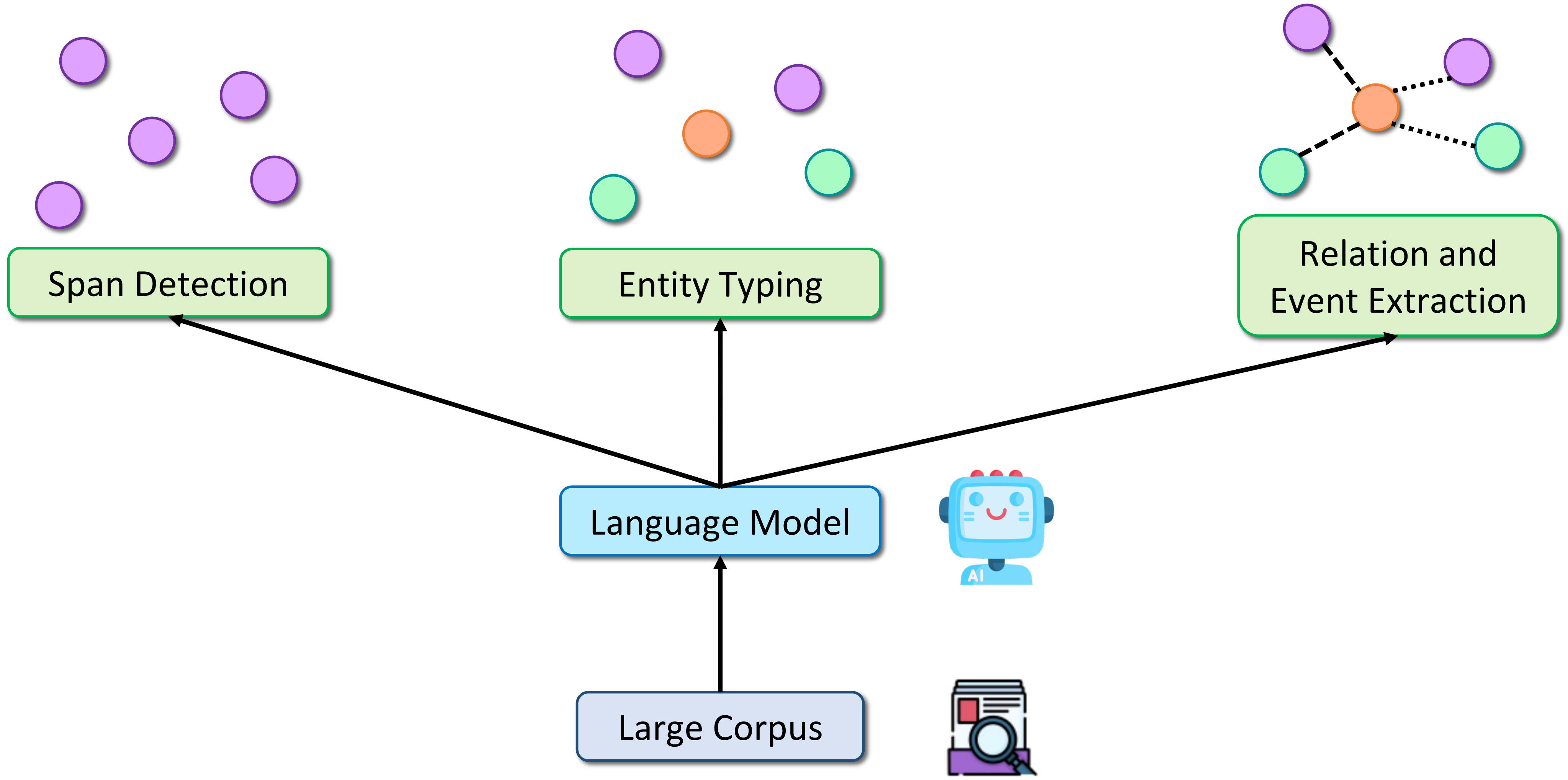
Computer Science, University of Illinois Urbana-Champaign

Aug 9, 2023

Tutorial Website:

Knowledge Base Construction

- Steps of Knowledge Base Construction



Outline

- Span Detection
  - Phrase Mining
  - Constituency Parsing
- Entity Typing
- Relation and Event Extraction

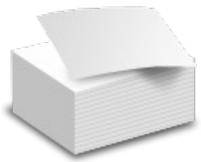
Why Phrase Mining?

- Identifying and understanding quality phrases from context is a fundamental task in text mining.

Scientific Papers

Expected Results
data mining machine learning information retrieval … support vector machine …

News Articles



Expected Results
US President Anderson Cooper Barack Obama … Obama administration …

- Quality phrases refer to informative multi-word sequences that "*appear consecutively in the text, forming a complete semantic unit in certain contexts or the given document*" [1].

[1] Geoffrey Finch. 2016. Linguistic terms and concepts. Macmillan International Higher Education

Why Phrase Mining?

w/o phrase mining

- ❑ What's "United"?
- ❑ Who's "Dao"?

w/ phrase mining

- ❑ United Airline!
- ❑ David Dao!

- ❑ Applications in NLP, IR, Text Mining
  - ❑ Text Classification
  - ❑ Indexing in search engine
  - ❑ Keyphrases for topic modeling
  - ❑ Text Summarization

Outline

- Span Detection
  - Phrase Mining
    - UCPhrase: Unsupervised Context-aware Quality Phrase Tagging [KDD'21]
  - Constituency Parsing
- Entity Typing
- Relation and Event Extraction

Previous Phrase Mining/Chunking Models

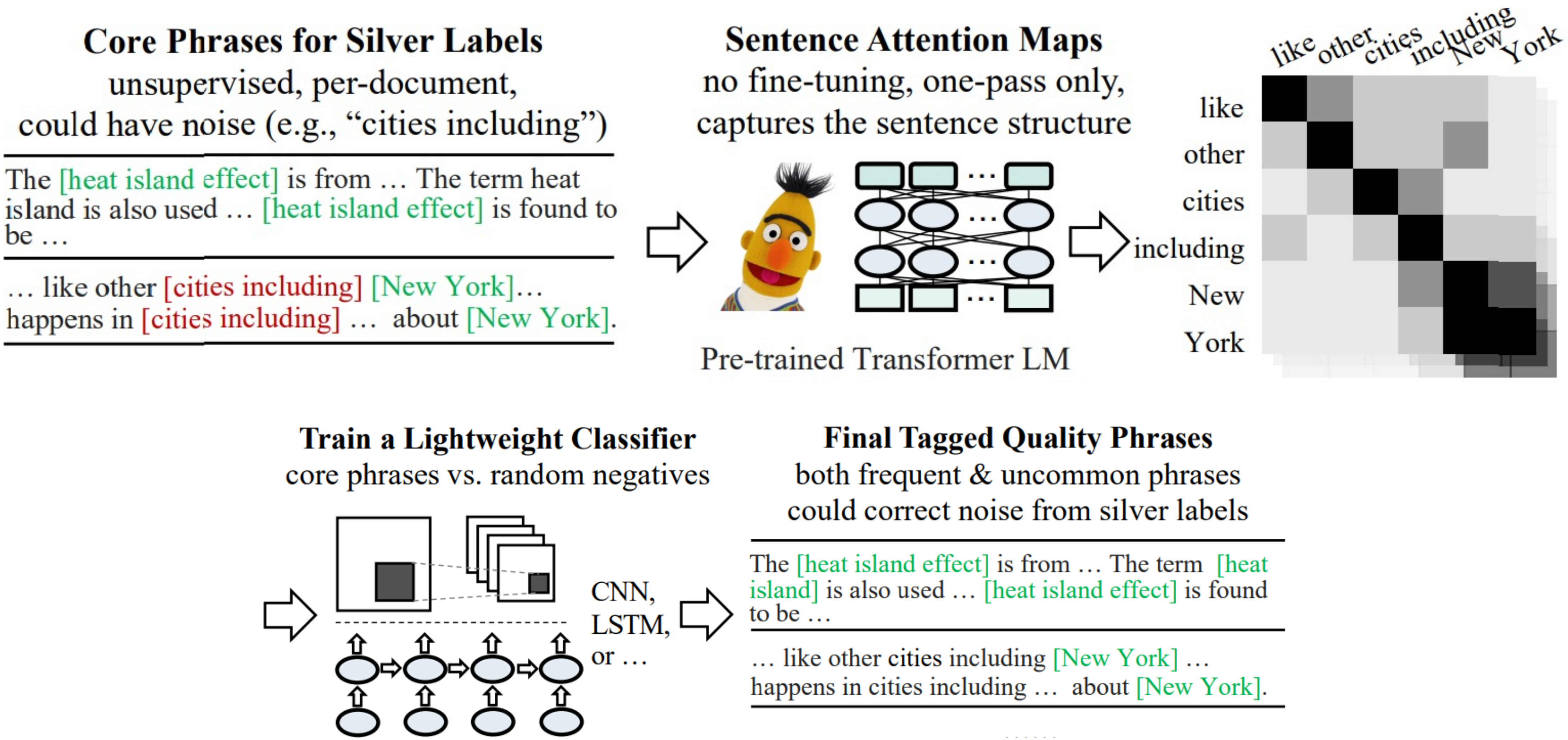
- Statistics-based models (*TopMine, SegPhrase, AutoPhrase)*
  - only work for frequent phrases, ignore valuable **infrequent / emerging phrases**
- Tagging-based models (*Spacy, StanfordNLP*)
  - do not have requirements for frequency
  - require **expensive and unscalable** sentence-level annotations for model training

Different Types of Supervisions

- Supervision
  - Human annotation
    - expensive, **hard to scale** to larger corpora and new domains
  - Distant supervision
    - tend to produce **incomplete labels** due to context-agnostic matching
      - e.g. "Heat [island effect] is found to be ..."
      - e.g. "Biomedical [data mining] is an important task where ..."
    - tend to match popular phrases, which form a small seen phrase vocabulary
      - easy for an embedding-based system to **memorize / overfit**

Framework of UCPhrase

- Silver Label Generation + Attention Map-based Span Prediction



Silver Label Generation

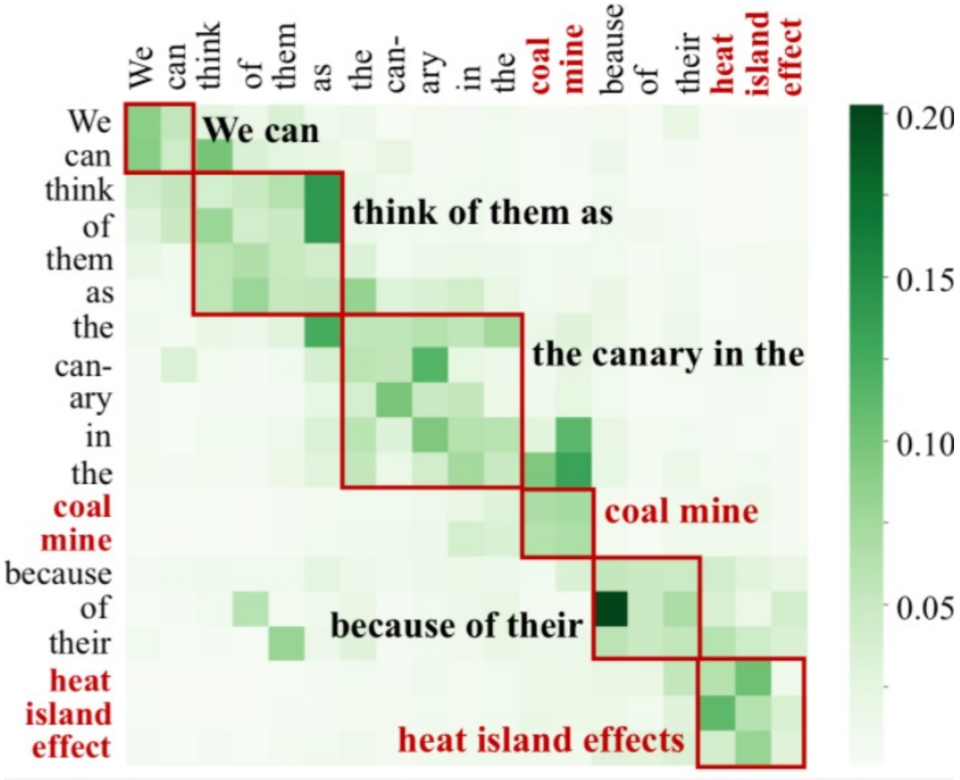
- How do human readers accumulate new phrases?
  - even without any prior knowledge we can recognize these consistently used patterns from a document
  - e.g., *task name, method name, dataset name, concepts* in a publication
  - e.g., *human name, organization, locations* in a news article
- Mining core phrases as silver labels
  - independently mine **max word sequential patterns** within each document
  - with each document as context
    - preserve contextual completeness ("biomedical data mining" vs. "data mining")
    - avoid potential noises from propagating to the entire corpus

Surface-Agnostic Feature Generation

- What's wrong with traditional embedding-based features?
  - embedding features are word identifiable -- it tells you which word you are looking at
  - easy to rigidly memorize all seen phrases / words in the training set / dictionary
  - fail to generalize to unseen phrases
- Good features for phrase recognition should be
  - agnostic to word **surface names** (so the model cannot rely on rigid memorization)
  - reveal the role that the span plays in the entire sentence (look at **sentence structure** rather than phrase names)

Attention Map

- Extract knowledge directly from a pre-trained language model
  - the **attention map** of a sentence vividly visualizes its **inner structure**
  - high quality phrases should have **distinct attention patterns** from ordinary spans



Phrase Tagging as Image Classification

- Viewing the generated feature as a 144-channel image of size K*K
  - train a lightweight 2-layer CNN model for binary classification: is a phrase or not
  - why CNN: capture word interactions (attentions) from various ranges, also fast for training and inference
- Efficient implementation
  - only train the CNN module, without fine-tuning LM

Quantitative Evaluation

Table 2: Evaluation results (%) of three tasks for all compared methods on datasets on two domains.

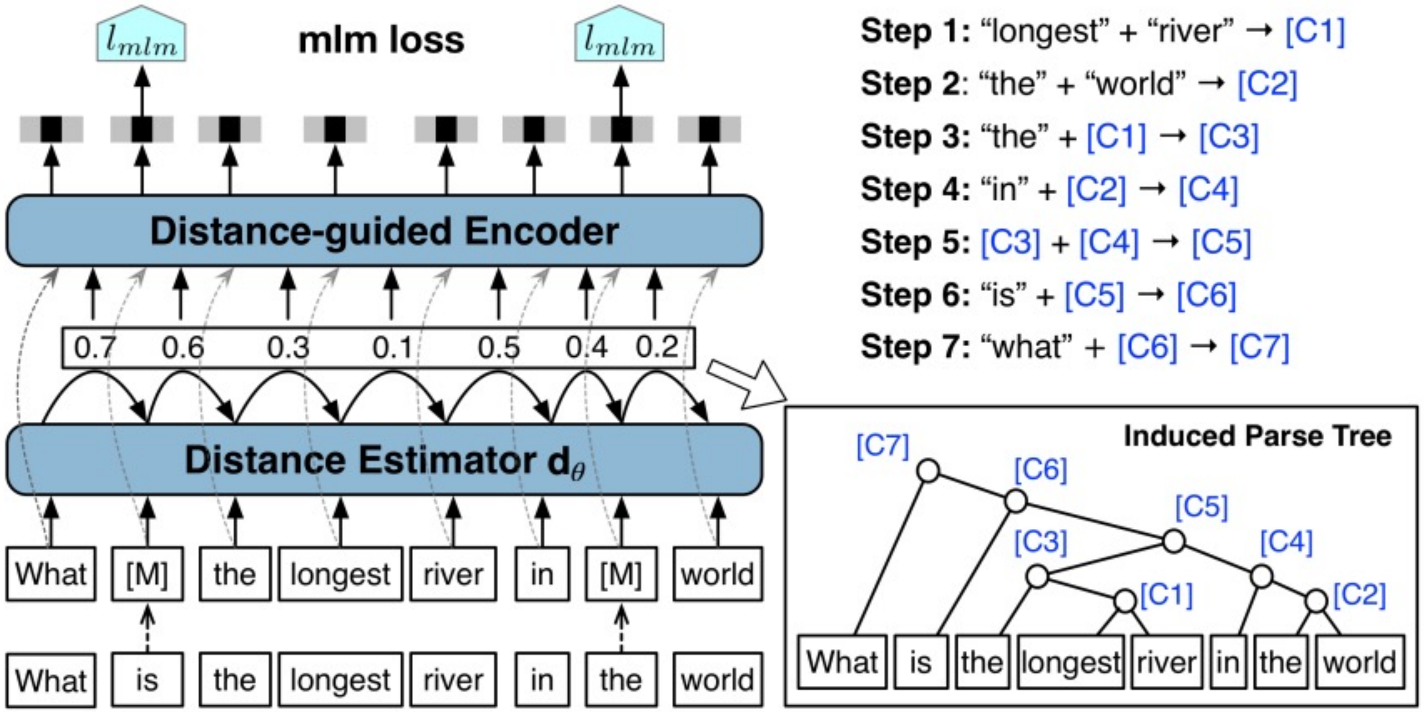
Method Type	Method Name	Task I: Phrase Ranking				Task II: KP Extract.				Task III: Phrase Tagging					
		KP20k		KPTimes		KP20K		KPTimes		KP20k			KPTimes		
		P@5K	P@50K	P@5K	P@50K	Rec.	F_1@10	Rec.	F_1@10	Prec.	Rec.	F_1	Prec.	Rec.	F_1
Pre-trained	PKE [3]	–	–	–	–	57.1	12.6	61.9	4.4	54.1	63.9	58.6	56.1	62.2	59.0
	Spacy [16]	–	–	–	–	59.5	15.3	60.8	8.6	56.3	68.7	61.9	61.9	62.9	62.4
	StanfordNLP [26]	–	–	–	–	51.7	13.9	60.8	8.7	48.3	60.7	53.8	56.9	60.3	58.6
Distantly Supervised	AutoPhrase [33]	97.5	96.0	96.5	95.5	62.9	18.2	77.8	10.3	55.2	45.2	49.7	44.2	47.7	45.9
	Wiki+RoBERTa	**100.0**	**98.5**	**99.0**	**96.5**	**73.0**	19.2	64.5	9.4	58.1	64.2	61.0	60.9	65.6	63.2
Unsupervised	TopMine [8]	81.5	78.0	85.5	71.0	53.3	15.0	63.4	8.5	39.8	41.4	40.6	32.0	36.3	34.0
	UCPhrase (ours)	96.5	96.5	96.5	95.5	72.9	**19.7**	**83.4**	**10.9**	**69.9**	**78.3**	**73.9**	**69.1**	**78.9**	**73.5**

Outline

- Span Detection
  - Phrase Mining
  - Constituency Parsing
    - Phrase-aware Unsupervised Constituency Parsing [ACL'2022]
- Named Entity Recognition
- Relation and Event Extraction

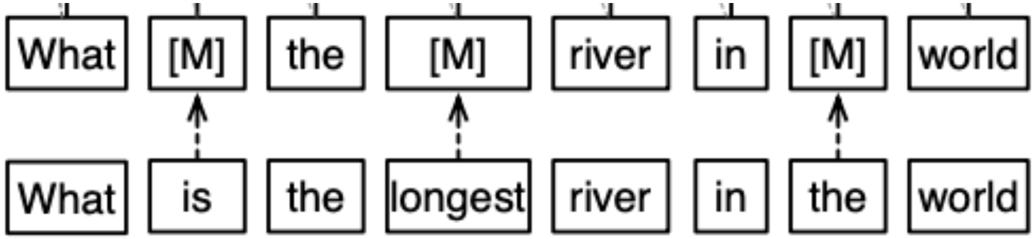
LM-based Unsupervised Constituency Parsing

- Represent discrete parsing tree as a distance sequence (given by a distance estimator)
- Distance information helps inject the parsing tree structure into encoder training via the MLM loss



Challenges With Current LM-Based Methods

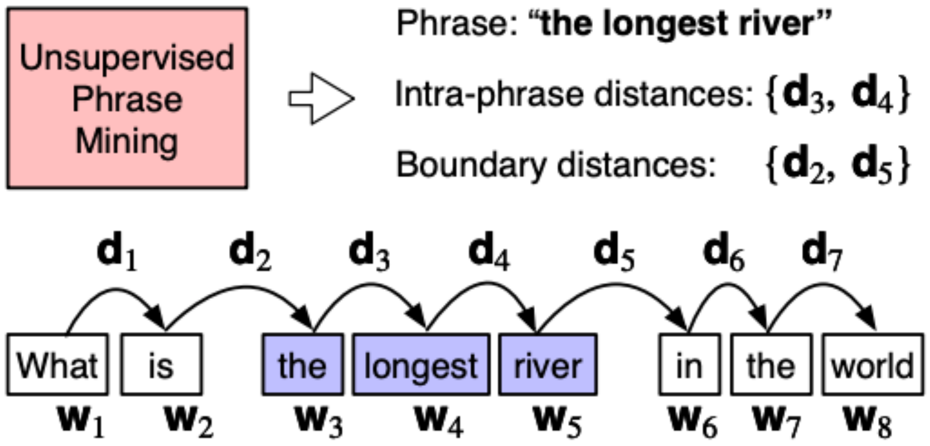
- The distance estimator is randomly initialized
  - yield suboptimal information for the encoder **in the cold start phase**
  - lead to suboptimal parsing accuracy due to **error accumulation**
- The token reconstruction task (MLM) mainly relies on the aggregation of **local information,** thus can hardly guide the model to manage **high-level structures across long distances**
  - Example: The prediction of "longest" mainly depends on its neighbor "river"



Phrase-Regularized Warm-Up

- Warm up the distance estimator via unsupervised extracted phrases
  - Can use any phrase tagger (e.g., UCPhrase)
- Encourage the average intra-phrase distance to be smaller than the average phrase boundary distance through a margin loss

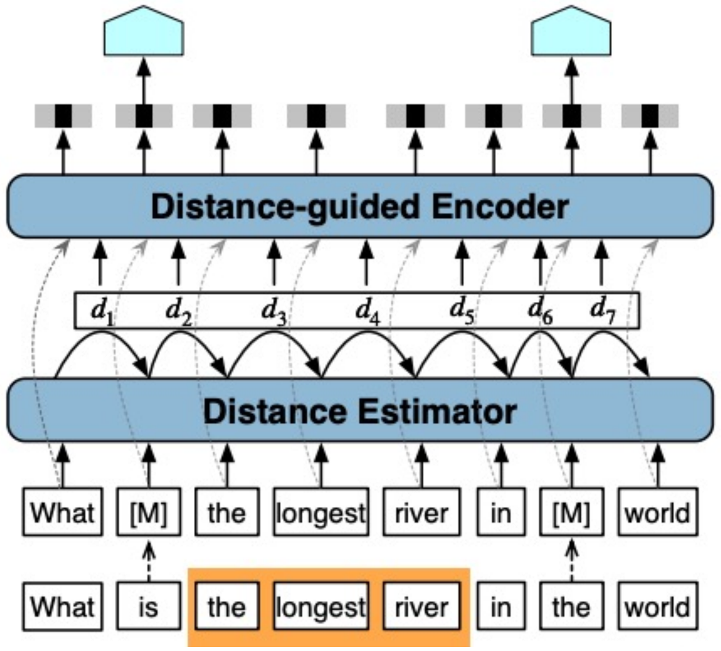
$$\ell_{phrase} = \frac{1}{4} \cdot (max(0, \mathbf{d}_3 - \mathbf{d}_2) + max(0, \mathbf{d}_3 - \mathbf{d}_5) + max(0, \mathbf{d}_4 - \mathbf{d}_2) + max(0, \mathbf{d}_4 - \mathbf{d}_5))$$



Phrase-aware Unsupervised Constituency Parsing [ACL'2022]

Phrase-Guided Masked Language Modeling

- Given a sentence with tagged local phrases, sample a subset of the phrases to be excluded from being masked out
- By doing so, we try to push the model out of its comfort zone of local structure learning, and encourage it to focus more on how the local constituents are connected



Results

- Phrase-guided masked language modeling (PMLM) and phrase-regularized warm-up (PRW) both help improve the performance of existing LM-based parsers

Methods	F1 (%)
PRPN (Shen et al., 2018a)	37.4
ON-LSTM (Shen et al., 2018b)	47.7
URNNG (Kim et al., 2019c)	52.4
C-PCFG (Kim et al., 2019b)	55.2
Neural L-PCFGs (Zhu et al., 2020)	55.3
TreeTransformer (Wang et al., 2019)	47.9
+ PMLM	<u>48.7</u>
+ PRW	<u>49.0</u>
+ PRW + PMLM	<u>49.3</u>
StructFormer (Shen et al., 2020)	54.0
+ PMLM	<u>54.1</u>
+ PRW	<u>55.3</u>
+ PRW + PMLM	<u>55.7</u>

Table 1: Unlabeled F1 score (%) for unsupervised constituency parsing on WSJ test set.

Method	NP	VP	ADJ	ADV	SBA	PP
PRPN	59.2	46.7	44.3	32.8	50.0	57.2
ON-LSTM	64.5	41.0	38.1	31.6	52.5	54.4
C-PCFG	74.7	41.7	40.4	52.5	56.1	68.8
TreeTransformer	63.7	37.1	32.3	56.8	37.0	49.7
+ PMLM	63.5	<u>37.9</u>	31.7	56.8	<u>38.0</u>	<u>50.4</u>
+ PRW	<u>64.2</u>	36.3	27.9	53.8	36.2	<u>53.0</u>
+ PRW + PMLM	<u>64.2</u>	<u>37.2</u>	29.6	53.7	35.9	<u>53.3</u>
StructFormer	73.7	43.2	53.4	70.5	51.8	64.5
+ PMLM	73.6	<u>43.7</u>	53.4	69.3	<u>51.9</u>	<u>64.6</u>
+ PRW	<u>74.0</u>	<u>44.9</u>	52.9	69.9	<u>52.7</u>	<u>69.4</u>
+ PRW + PMLM	<u>74.2</u>	<u>45.1</u>	53.2	69.3	<u>53.9</u>	<u>70.1</u>

Table 2: Recall scores (%) of typed gold constituents.

Outline

- Span Detection
- Entity Typing
  - Few-shot Entity Typing
    - Automatic Label Interpretation and Generating New Instance for Entity typing [KDD’22]
  - Zero-shot Entity Typing
- Relation and Event Extraction

Motivation

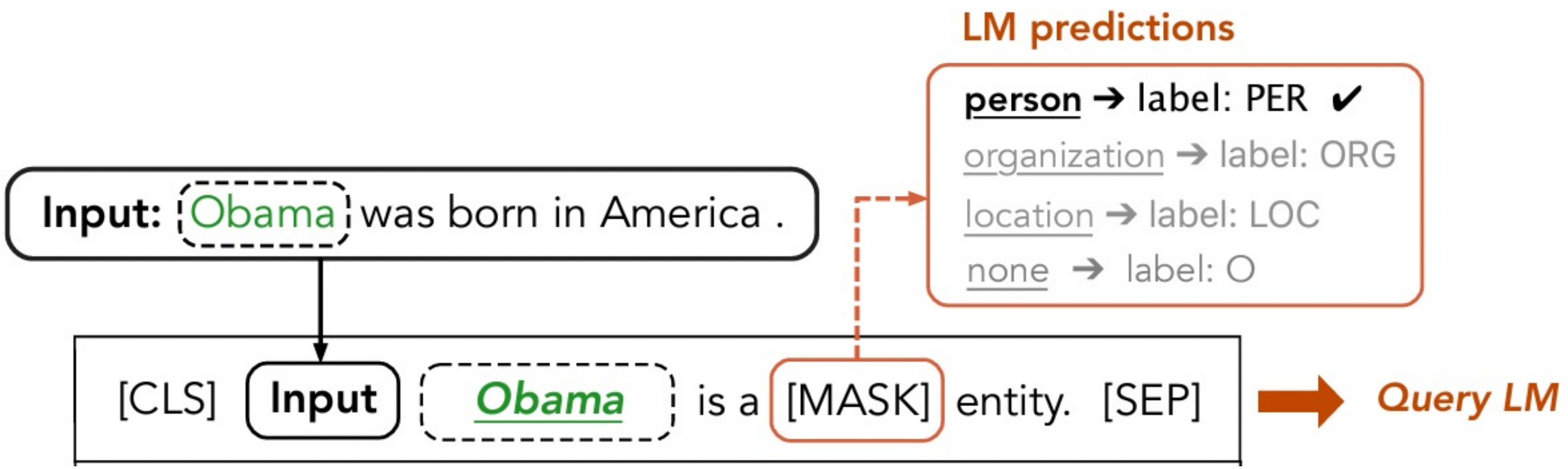
- Entity typing is a fundamental task in text mining with a wide spectrum of applications
  - question answering
  - knowledge base construction
  - dialog systems
  - …
- Deep neural models have achieved enormous success for entity typing
- However, a common bottleneck of training deep learning models is the acquisition of abundant high-quality human annotations (every entity in the sequence needs to be labeled!)

Few-shot Entity Typing

- Current entity typing models are trained for a series of fixed categories (e.g., PERSON, LOCATION, etc.) using large amounts of labeled data.
- Few-shot entity typing learns to transfer to new domains/categories with only a few training examples.

LM-based Entity Typing

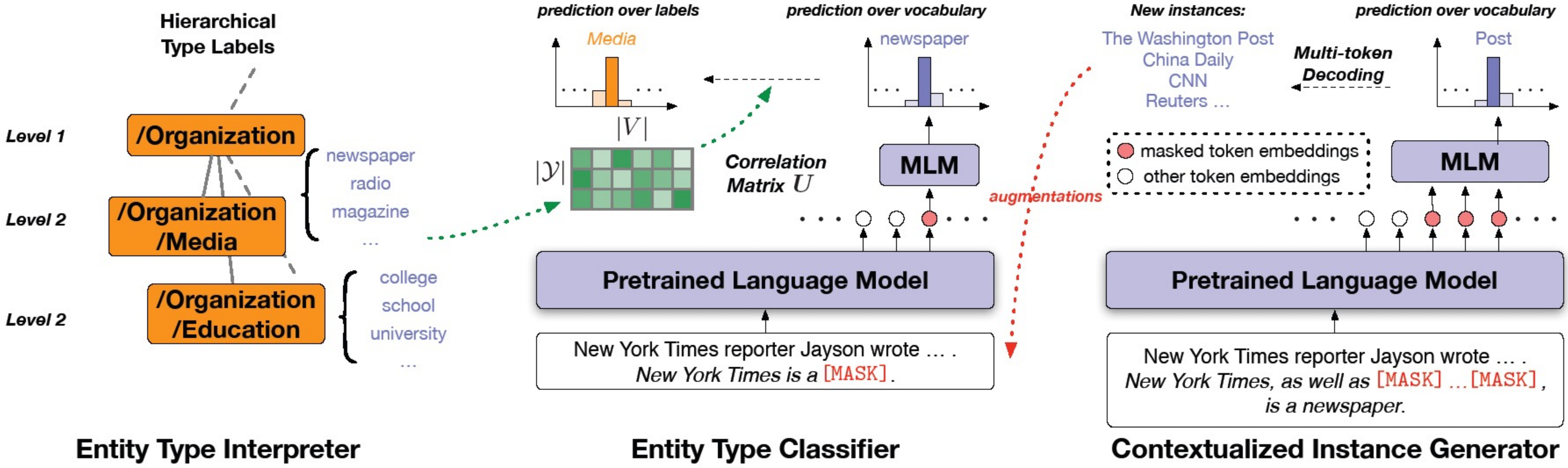
- An example of prompting language models for named entity recognition.



Limitations of current pipeline

- A verbalizer needs to be manually created for mapping words to labels (e.g., person/actor -> PER, organization/company -> ORG)
- Current approaches have not fully utilized the power of LMs
  - **representation** models that predict entity types based on entity instance representations
  - the **generation** power of LMs acquired through extensive general-domain pretraining can be exploited to generate new entity instances
  - model can be trained with more instances for better generalization

Overall Framework of ALIGNIE (Automatic Label Interpretation and Generating New Instance for Entity typing)



(Left): With a given type label hierarchy, an entity type interpretation module relates all the words in the vocabulary with the label hierarchy by a correlation matrix.

(Middle): An entity typing classifier maps the word probability at the [MASK] position to type probability using the correlation matrix.

(Right): A type-based contextualized instance generator uses an entity mention and its predicted type to construct a template for new instance generation to augment the training set.

PLM-based Instance Generator

- E.g., a *newspaper* entity “New York Times” ➡ more newspaper names

Generation Template :

[Context]. **New York Times**, as well as [MASK] [MASK] [MASK], is a *newspaper*.

Entity Mention

ranges from
1 to the length of
original entity mention

Predicted by
Entity Type
Classifier

Multi-Token Instance Generation

- We randomly choose one [MASK] token at each step, and sample from its output token probability to fill in a word.

E.g.

New York Times, as well as the$_1$ [MASK] [MASK] is a newspaper.
New York Times, as well as the$_1$ Washington$_2$ [MASK] is a newspaper.
New York Times, as well as the$_1$ Washington$_2$ Post$_3$ is a newspaper.

$$\text{Score}(\widetilde{\boldsymbol{m}}) = \sum_{i=1}^{|\widetilde{\boldsymbol{m}}|} \log(s_i)$$

The conditional probability at each step

The next blank to be filled in is randomly selected, therefore the order is not always from left to right.

Generated New instances based on predicted types of example entities

- Multi-token instances

Generation from **multi-token** entities		
Context & **entity mention**	MLM predicted type	Generated new instances
The album also included the song "Vivir Lo Nuestro," a duet with **Marc Anthony**.	singer	Beyonce, Jennifer Lopez, Rihanna, Taylor Swift, Lady Gaga, Michael Jackson, ...
The film was released on August 9, 1925, by **Universal Pictures**.	company	Warner Brothers, Paramount Pictures , Columbia Pictures, Lucasfilm, Hollywood Pictures, ...
Everland hosted 7.5 million guests in 2006, ranking it fourth in Asia behind the two **Tokyo Disney Resort** parks and Universal Studios Japan, while Lotte World attracted 5.5 million guests to land in fifth place.	park	Lotte World, Universal Studios Japan, Shanghai Disney World , Orlando Universal Studios, ...
The site of Drake's landing as officially recognised by the **U.S. Department of the Interior** and other agencies is Drake's Cove.	government agency	the Department of Homeland Security, the Bureau of Land Management, the Federal Bureau of Investigation, the United States Forest Service, the National Institutes of Health, ...
Pikmin also make a cameo during the process of transferring downloadable content from a **Nintendo DSi** to a 3DS, with various types of Pikmin carrying the data over.	handheld	3DS, 2DS, Wii U, Nintendo Switch, the PSP, PlayStation Vita, ...

Main Results

Method	OntoNotes			BBN			Few-NERD		
	(Acc.)	(Micro-F1)	(Macro-F1)	(Acc.)	(Micro-F1)	(Macro-F1)	(Acc.)	(Micro-F1)	(Macro-F1)
5-Shot Setting									
Fine-tuning	28.60	50.70	51.60	51.03	60.03	58.22	36.09	48.56	48.56
Prompt-based MLM	32.62	60.97	61.82	67.00	75.23	73.55	44.69	59.24	59.24
PLET	48.57	70.63	75.43	71.23	79.22	78.93	56.94	68.81	68.81
ALIGNIE (- hierarchical reg.)	52.74	**77.55**	79.72	72.15	80.35	80.40	59.01	70.91	70.91
ALIGNIE (- new instances)	51.10	72.91	76.88	73.50	81.62	81.31	57.41	69.47	69.47
ALIGNIE	**53.37**	77.21	**80.68**	**75.44**	**82.20**	**82.30**	**59.72**	**71.90**	**71.90**
Fully Supervised Setting									
Fine-tuning	56.70	75.21	78.86	78.06	82.39	82.60	79.75	85.74	85.74
Prompt-based MLM	55.18	74.57	77.47	77.10	81.77	82.05	77.38	85.22	85.22

- Prompt-based results have higher performance than vanilla fine-tuning in few-shot settings. In fully supervised settings, however, fine-tuning performs a little better than prompt-based MLM.
- ALIGNIE can even outperform fully supervised setting on OntoNotes and BBN, but cannot on Few-NERD. This is because the training set of OntoNotes and BBN are automatically inferred from external knowledge bases, and can contain much noise.

Outline

- ❑ Span Detection
- ❑ Entity Typing
  - ❑ Few-shot Entity Typing
  - ❑ Zero-shot Entity Typing
    - ❑ ONTOTYPE: Ontology-Guided Annotation-Free Fine-Grained Entity Typing
- ❑ Relation and Event Extraction

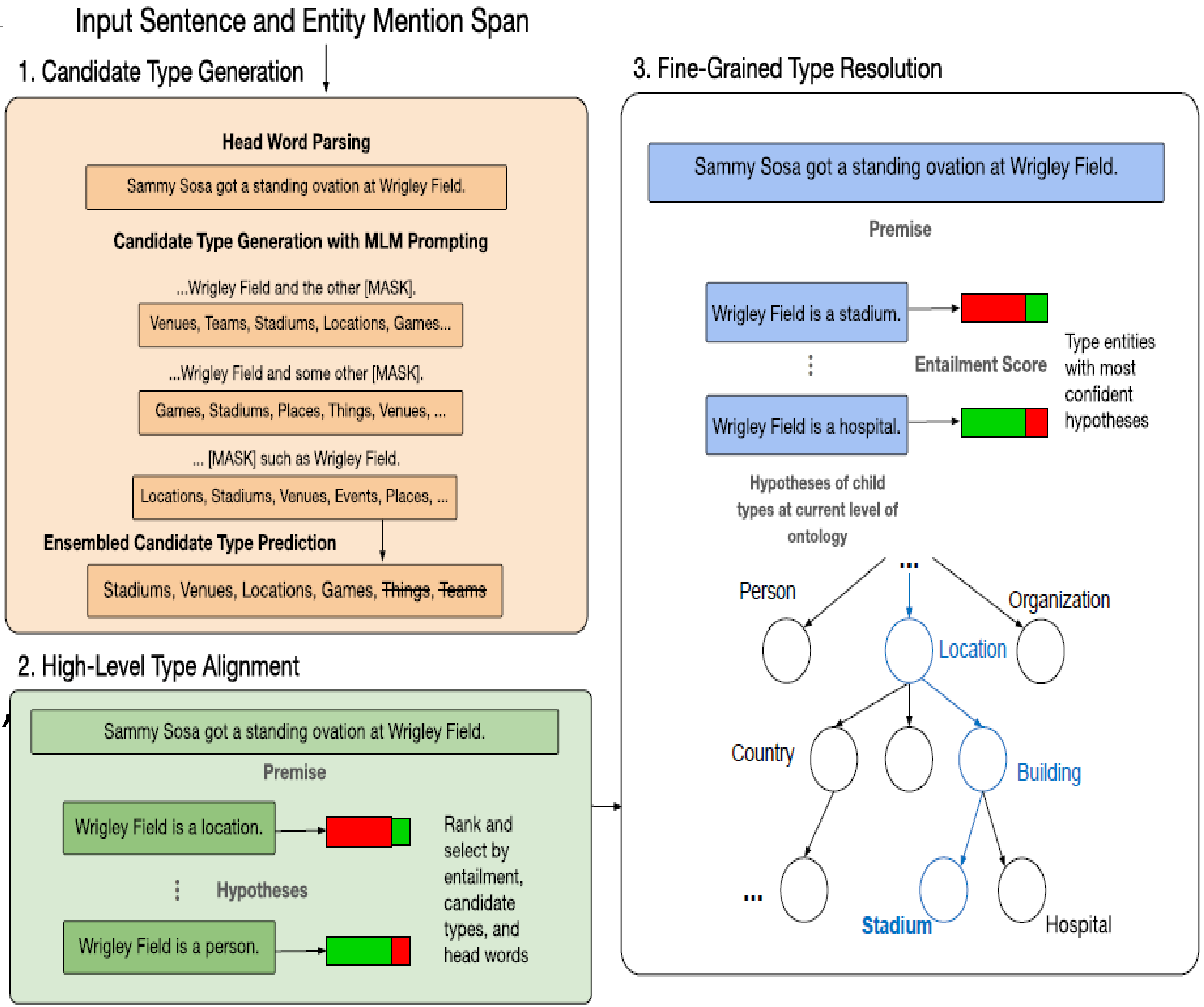
OntoType: Ontology-Guided Entity Typing

- Zero-shot entity typing: Assigns fine-grained semantic types to entities without any annotations
  - Ex. *Sammy Sosa* [Person/Player] got a standing ovation at *Wrigley Field* [Location/Building/Stadium]
- Challenges of weak supervision based on masked language model (MLM) prompting
  - A prompt generates a set of tokens, some likely vague or inaccurate, leading to erroneous typing
  - Not incorporate the rich structural information in a given, fine-grained type ontology
- OntoType: Ontology-guided, Annotation-Free, Fine-Grained Entity Typing
  - Ensemble multiple MLM prompting results to generate a set of type candidates
  - Progressively refine type resolution, from coarse to fine, following the type ontology, under the local context with a natural language inference model
- OntoType: Outperforms the SOTA zero-shot fine-grained entity typing methods

Tanay, Komarlu, et al., “ONTOTYPE: Ontology-Guided Annotation-Free Fine-Grained Entity Typing”, submitted for publication, 2022

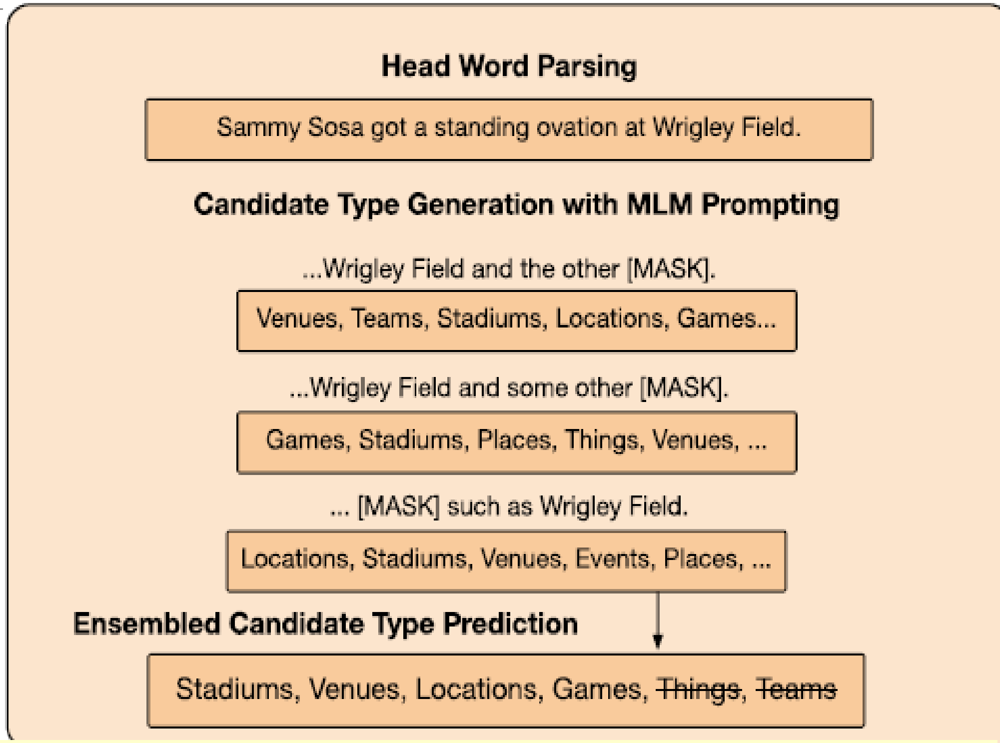
Overall Framework of OntoType : Three Steps

- Candidate type generation
  - Candidate type generation with multiple MLM prompting
  - Ensembled candidate type prediction
  - Ex. Stadium, venue, location, games, ~~things, teams~~
- High-level type alignment by entailment (local context + NLI)
- Progressively refine type resolution, from coarse to fine, following the type ontology
- Type ontology used at every step



OntoType: Step 1: Candidate Type Generation

- Head Word Parsing
  - Mention's head word in the input text is often the cue that explicitly matches a mention to its type
  - Ex. "Governor Arnold Schwarzenegger gives a speech …"
  - Use the Stanford Dependency Parser to extract head word
  - Leverage the head words of the input entity to select an initial context-sensitive coarse-grained type
- Ensembled MLM Prompting
  - Leverage a BERT masked language model and Hearst patterns to generate candidate types for the target mentions
  - Ensemble n patterns to generate the best candidate types
  - Consolidated candidates that are generated by a majority of the n Hearst patterns
    - Ex. For e_1, "Stadiums, Venues, Locations, Games" retain but the noisy types "Things" and "Teams" are removed

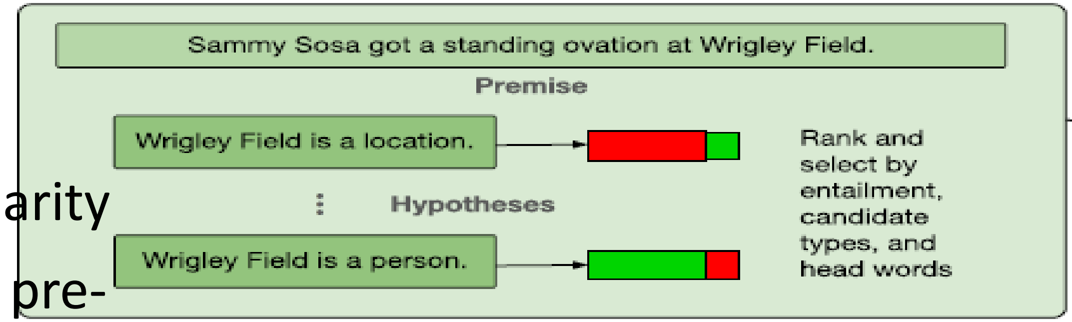


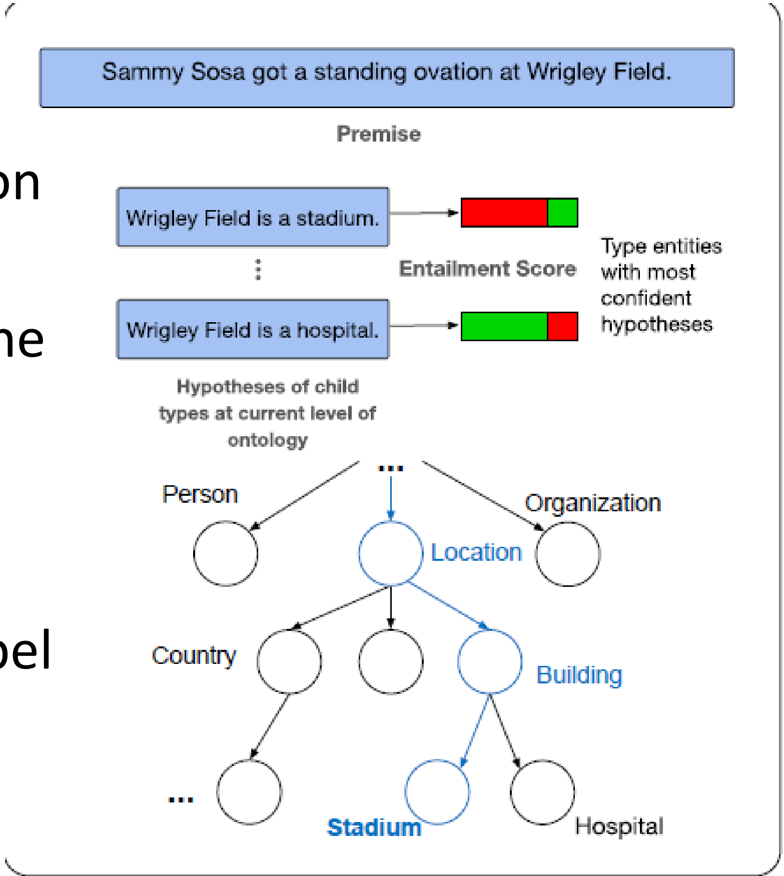
Four Hearst Patterns give the highest quality hypernyms with simple type mapping on the Ontonotes dataset

Hearst Pattern	Prec	Rec	F1
[MASK] such as	53.3	72.4	61.4
such [MASK] as	47.9	68.7	56.5
and some other [MASK]	48.8	66.6	56.4
and the other [MASK]	47.6	68.3	56.1

OntoType: Steps 2 & 3: High-Level Type Resolution & Progressive Type Refinement

- High-level type alignment by entailment
  - Align generated candidate types to several high-level types in the type ontology by Word2Vec+ cosine similarity
  - Then select the most accurate high-level types with a pre-trained entailment language model (NLI)
  - Ex. "Locations", "Stadiums", "Venues", and "Games" are closely related to the high-level type “Location"; NLI further ranks Location over Person and Organization
- Progressively refine type resolution, from coarse to fine, following the type ontology
  - Ex. At the 2nd level of ontology, it generates the hypotheses and ranks all child types of "location“
    - This consolidates and selects "building" as the highest ranked label
    - At a deeper level, it selects the final type "stadium”
- Type ontology is used at every step





OntoType: Performance Study

- Use 3 benchmark FET datasets: NYT, Ontonotes, and FIGER:

Datasets	Ontonotes	FIGER	NYT
# of Types	89	113	125
# of Documents	300k	3.1M	295k
# of Entity Mentions	242K	2.7M	1.18M
# of Train Mentions	223K	2.69M	701K
# of Test Mentions	8,963	563	1,010

Model	Prec	Rec	Ma-F1
OntoType$_{BERT}$	82.3	77.1	79.6
OntoType$_{RoBERTa}$	81.9	76.9	79.4
OntoType$_{Word2Vec}$	**84.7**	**78.4**	**81.5**

Compare with Zoe on Ontonotes with modified ontology

Model	Acc	Mi-F1	Ma-F1
Zoe	57.1	70.7	73.4
OntoType + Modified Ontology	**58.9**	**71.1**	**78.7**

- Compare with supervised and 0-shot methods:

Settings	Model	NYT			FIGER			Ontonotes		
		Acc	Mi-F1	Ma-F1	Acc	Mi-F1	Ma-F1	Acc	Mi-F1	Ma-F1
Supervised	AFET [16]	-	-	-	55.3	66.4	69.3	55.1	64.7	71.1
	UFET [2]	-	-	-	-	-	-	59.5	71.8	76.8
	BERT-MLMET [3]	-	-	-	-	-	-	67.44	80.35	85.44
Zero-Shot	ZOE [25]	62.1	73.7	76.9	**58.8**	**71.3**	74.8	50.7	60.8	66.9
	OTyper [22]	46.4	65.7	67.3	47.2	67.2	69.1	31.8	36.0	39.1
	DZET [14]	27.3	53.1	51.6	28.5	56.0	55.1	23.1	28.1	27.6
	MZET [23]	30.7	58.2	56.7	31.9	57.9	55.5	33.7	43.7	42.3
	OntoType + Original Ontology (Ours)	-	-	-	49.1	67.4	75.1	**65.7**	**73.4**	**81.5**
	OntoType + Modified Ontology (Ours)	**69.6**	**78.4**	**82.8**	51.1	68.9	**77.2**	-	-	-

OntoType: Case Study

MZET	**US President Joe Biden** \Person\Politician was one of many foreign leaders to speak with **President Zelensky** \Person\Politician, and he "pledged to continue providing **Ukraine** \Location with the support needed to defend itself, including advanced air defence systems", **the White House** \Location\Building said.	Trailing two games to one in **the NBA Finals** \Other\Event and facing the daunting task of trying to beat **the Boston Celtics** \Organization\Company in the hostile environment of **TD Garden** \Location\Building on Friday night, the Warriors knew they needed to summon one of the best efforts of their dynastic run in order to even the best-of-seven series.
ZOE	**US President Joe Biden** \Person\Politician was one of many foreign leaders to speak with **President Zelensky** \Person\Politician, and he "pledged to continue providing **Ukraine** \Location\Country with the support needed to defend itself, including advanced air defence systems", the **White House** \Location\Building said.	Trailing two games to one in **the NBA Finals** \Other\Event and facing the daunting task of trying to beat **the Boston Celtics** \Organization\Sports_Team in the hostile environment of **TD Garden** \Location\Building\Sports_Facility on Friday night, the Warriors knew they needed to summon one of the best efforts of their dynastic run in order to even the best-of-seven series.
ONTOTYPE	**US President Joe Biden** \Person\Politician\President was one of many foreign leaders to speak with **President Zelensky** \Person\Politician\President, and he "pledged to continue providing **Ukraine** \Location\Country with the support needed to defend itself, including advanced air defence systems", **the White House** \Organization\Government said.	Trailing two games to one in **the NBA Finals** \Other\Event\Finals and facing the daunting task of trying to beat **the Boston Celtics** \Organization\Sports_Team\Basketball_Team in the hostile environment of **TD Garden** \Location\Building\Sports_Facility on Friday night, the Warriors knew they needed to summon one of the best efforts of their dynastic run in order to even the best-of-seven series.

- See how different methods perform on news articles with a modified FIGER type ontology

Outline

- Span Detection
- Entity Typing
- Relation and Event Extraction
  - Relation Extraction
    - Document-Level Relation Extraction
    - Corpus-Level Relation Extraction
  - Event Discovery

Document-Level Relation Extraction

- Document-level relation extraction (DocRE)
  - Extract semantic relations among entity pairs in a document
- Blindly considering the full document?
  - A subset of the sentences in the doc (“evidence”) should often be sufficient to identify the relation
- An evidence-enhanced DocRE framework: EIDER
  - Efficiently extracts evidence and effectively leverages the extracted evidence to improve DocRE
- Using a document-level relationship extraction dataset DocRED (2019)
- Relation extraction benefits natural language understanding in many ways
  - Ex. Knowledge graph construction

Head:Hero of the Day Tail:the United States Rel:[country of origin]
GT evidence sentences: [1,10] Extracted evidence: [1,10]

Original document as input: [1] Load is the sixth studio album by the American heavy metal band Metallica, released on June 4, 1996 by Elektra Records in **the United States** … [9] It was certified 5×platinum … for shipping five million copies in **the United States**. [10] Four singles—"**Hero of the Day**", "Until It Sleeps", "Mama Said", and "King Nothing" — were released as part of the marketing campaign for the album.
Prediction scores: NA: 17.63 **country of origin**: 14.79

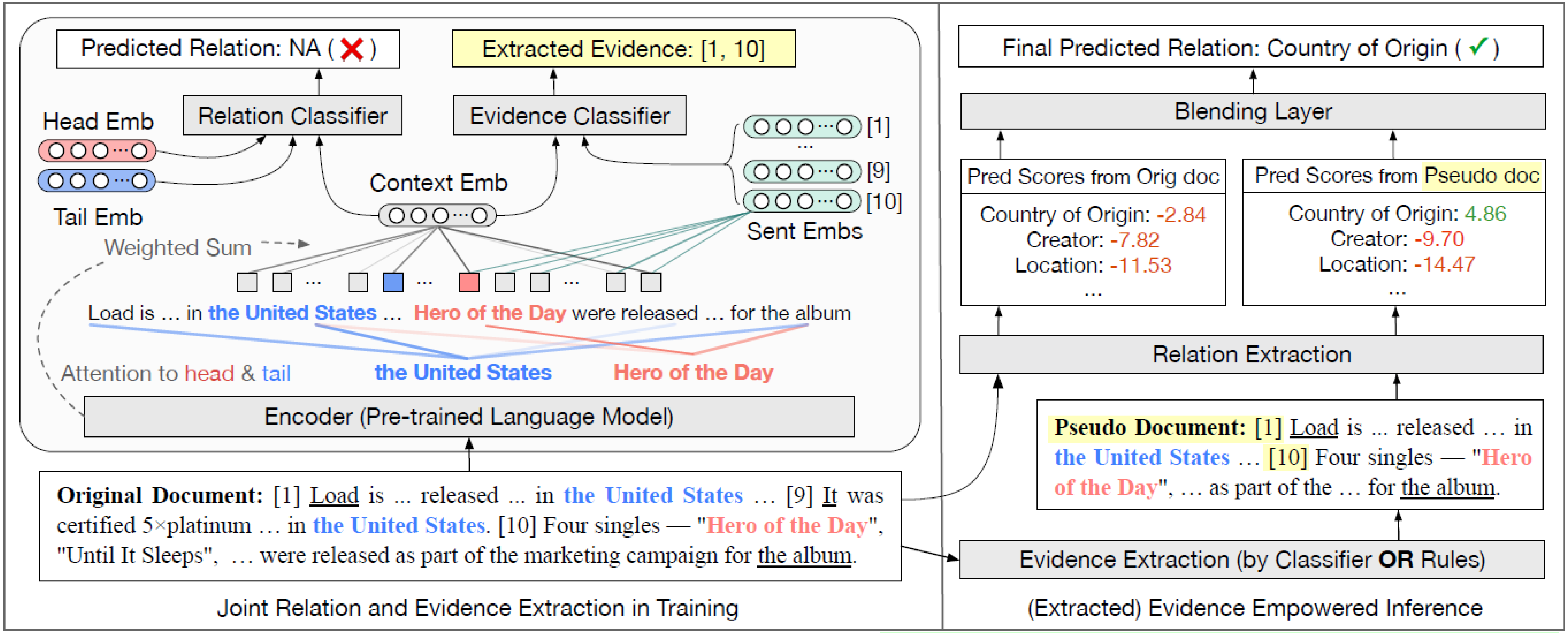
Extracted evidence as input: [1] Load is the sixth studio album … released … in **the United States** … [10] Four singles — "**Hero of the Day**", … were released … for the album.
Prediction scores: **country of origin**: 18.31 NA: 13.45

Final prediction of our model: **country of origin** (✓)

Only need [1]+[10] to identify [head, relation, tail]

Yiqing Xie, Jiaming Shen, Sha Li, Yuning Mao, Jiawei Han, “EIDER: Evidence-enhanced Document-level Relation Extraction”, ACL’22 Findings

EIDER Architecture



The left part (the training stage), we jointly extract relation and evidence using multi-task learning, where the two tasks have their own classifier and share the base encoder

The right part (the inference stage), we fuse the predictions on the original document and the extracted evidence using a blending layer

Corpus-Level Relation Extraction

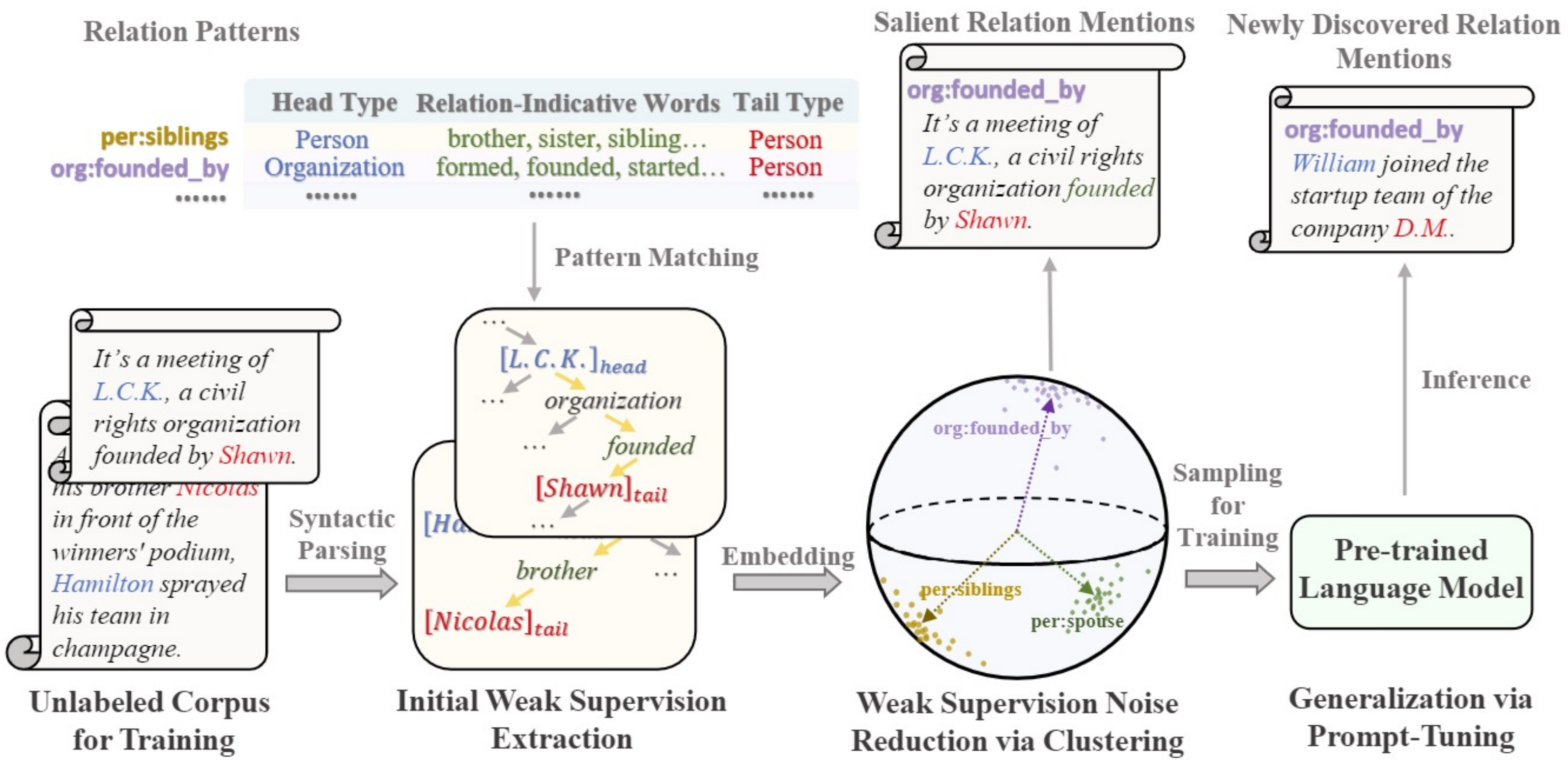


Fig. 2. Framework overview. Our model mainly consists of three steps: (1) relation triple representation extraction, (2) latent space clustering, and (3) prompt-tuning with sub-prompts.

Sizhe Zhou, Suyu Ge, Jiawei Han, "Corpus-Based Relation Extraction by Identifying and Refining Relation Patterns", ECMLPKDD'23

- Utilized additional representation of relation triple for initial weak supervision extraction and latent clustering for further denoising
- Applied further prompt tuning for context understanding and pattern generalization

Experiment Results

Model	TACRED				TACREV				ReTACRED			
K	4	8	16	Mean	4	8	16	Mean	4	8	16	Mean
w/ weak supervision												
EXACT MATCHING*	-	-	-	48.87	-	-	-	53.67	-	-	-	54.86
COSINE	23.28	26.60	37.16	29.01	21.43	30.85	41.21	31.16	28.12	35.00	44.54	35.89
COSINE*	-	-	-	58.88	-	-	-	60.80	-	-	-	68.59
RCLUS NOISY	45.35	50.94	55.73	50.67	50.41	61.67	**66.85**	59.64	56.89	65.81	71.09	64.60
RCLUS BALANCED	45.19	55.71	59.33	53.41	55.36	58.74	64.56	59.55	53.84	65.27	71.03	63.38
RCLUS	**49.89**	**56.65**	**60.26**	55.60	**56.94**	**63.75**	66.50	62.40	**61.03**	**68.78**	**72.23**	67.35
w/ ground truth supervision												
FINE-TUNING	13.62	26.09	32.07	23.93	18.75	25.21	35.12	26.36	17.36	31.77	42.63	30.59
GDPNET	13.79	28.42	43.11	28.44	15.61	24.59	42.12	27.44	19.20	35.79	52.84	35.94
PTR	39.16	49.46	54.67	47.76	47.18	51.58	59.17	52.64	51.27	62.60	71.11	61.66

Table 1. F_1 scores (%) on full test set with different sizes (K = 4, 8, 16) for each relation label.

Model	TACREV Precision	TACREV Recall	TACREV F_1
RCLUS			
w/ Weak	59.30	49.02	53.67
w/ Prompt	48.25	75.73	58.95
w/ Weak + Prompt	58.80	72.07	64.76
w/ Weak + Cluster	63.62	40.61	49.57
w/ Weak + Cluster + Prompt	60.76	74.29	66.85
w/ Weak + Cluster + Prompt*	57.85	78.47	66.61

Table 3. Ablation study of RCLUS

- Leading low-resource performance
- Each component is indispensable
  - Weak supervision provides relatively high recall
  - Clustering provides relatively high precision
  - Prompt-tuning is important for boosting recall

Outline

- Span Detection
- Entity Typing
- Relation and Event Extraction
  - Relation Extraction
  - Event Discovery
    - Argument Role Prediction
      - Open-Vocabulary Argument Role Prediction [EMNLP'22]
    - Event Chain Mining

Open-Vocabulary Argument Role Prediction

- Related Work:
  - Most of existing studies rely on hand-crafted ontologies (costly, cannot generalize)
  - A few studies try to automatically induce argument roles (limited pre-defined glossary)
- **New Task**: Infer a set of argument role names for a given event type to describe the crucial relations between the event type and its arguments

Event Type: Earthquake

The **2007 Peru earthquake**, which measured 8.0 on the moment magnitude scale, hit the central coast of Peru on August 15 at 23:40:57 UTC (18:40:57 local time) and lasted two minutes. The epicenter was located 150 km (93 mi) south-southeast of Lima at a depth of 39 km (24 mi). The United States Geological Survey National Earthquake Information Center reported that it had a maximum Mercalli intensity of IX. The Peruvian government stated that 519 people were killed by the quake.

Argument Role Prediction

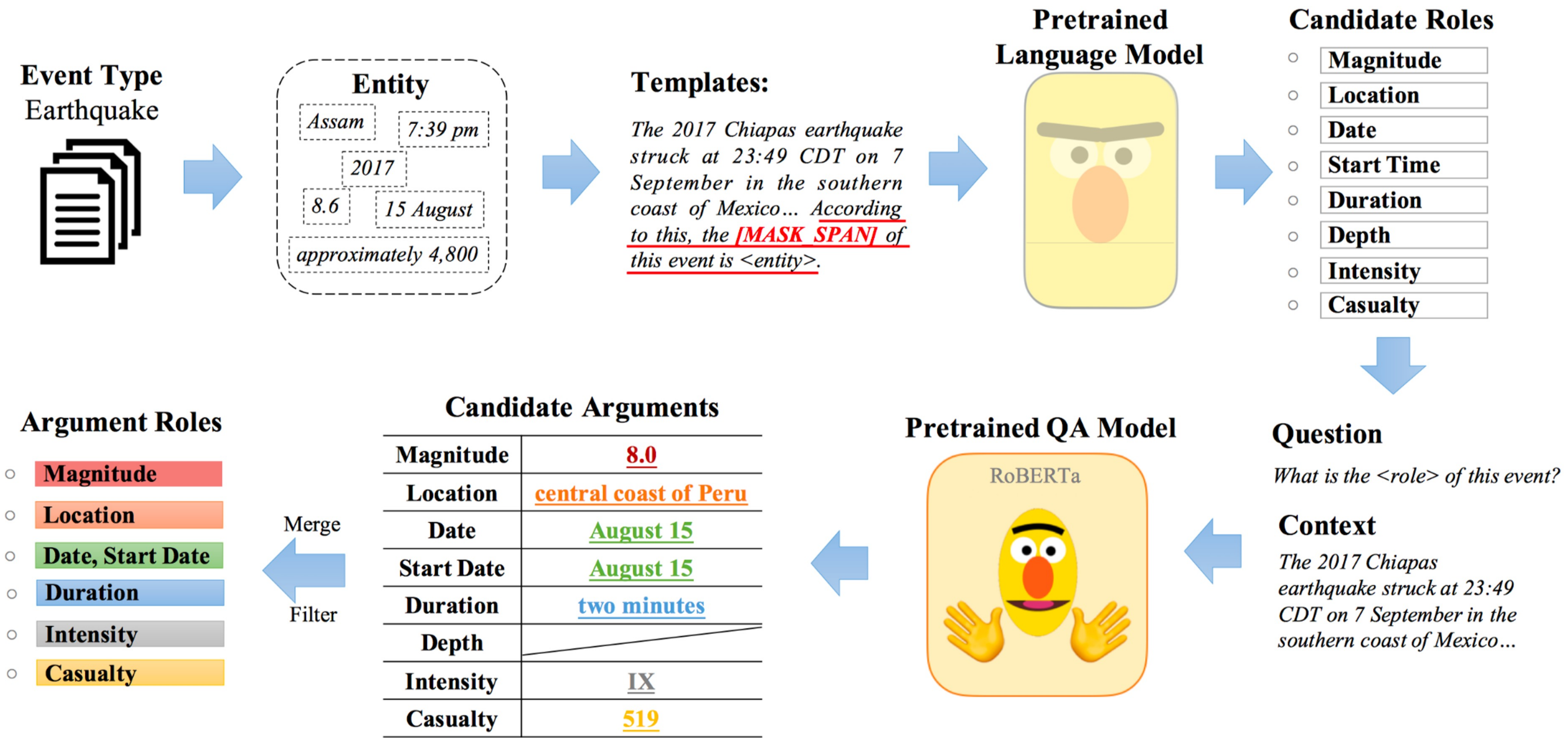
- Magnitude
- Location
- Date
- Time
- Duration
- Depth
- Intensity
- Casualty

Downstream Task
Argument extraction

Magnitude	8.0
Location	central coast of Peru
Date	August 15
Time	23:40:57 UTC
Duration	two minutes
Depth	39 km
Intensity	IX
Casualty	519

Yizhu Jiao, Sha Li, Yiqing Xie, Ming Zhong, Heng Ji and Jiawei Han "Open-Vocabulary Argument Role Prediction for Event Extraction", EMNLP'22

Framework for RolePred (Argument Role Prediction)



RolePred 1: Candidate Role Generation

- Predict candidate role names for named entities by casting it as a prompt-based in-filling task
  - Prompt Construction: (using Generation Model : T5)
    - *Context*. According to this, the ⟨MASK SPAN⟩ of this Event Type is Entity.
  - Ex. *The 1964 Alaskan earthquake, also known as the Great Alaskan earthquake, occurred at 5:36 PM AKST on Good Friday, March 27.* According to this, the ⟨MASK SPAN⟩ of this earthquake is 5:36 PM.
    - ⟨MASK SPAN⟩ is expected to be filled with *time* (or *start time*) as the argument role
  - Considering the entity's general semantic type: person, location, number, etc., we slightly alter the prompt to fluently and naturally support the unmasking argument roles

Prompt design for different entities

Entity Type	Prompt
PERSON	*According to this, Entity play the role of ⟨MASK SPAN⟩ in this Event Type.*
LOCATION	*According to this, the ⟨MASK SPAN⟩ is Entity in this Event Type.*
NUMBER	*According to this, the number of ⟨MASK SPAN⟩ of this Event Type is Entity.*
OTHER TYPES	*According to this, the ⟨MASK SPAN⟩ of this Event Type is Entity.*

RolePred 2: Candidate Argument Extraction

- Formulate the argument extraction problem into question-answering task
  - Input: follow a standard BERT-style format (Model: BERT based pretrained QA model)
    - [CLS] What is the Event Role in this Event Type event? [SEP] Document [SEP]
  - Ex. [CLS] What is the *casualty* in this *pandemic* event? [SEP] *The COVID-19 pandemic is an ongoing global pandemic of coronavirus disease. It's estimated that the worldwide total number of deaths has exceeded five million ...* [SEP]
    - The argument is expected to be five million
    - Note that, for some roles, a given document may not mention its argument. That is, the above-constructed question can be unanswerable. Thus, for each extracted answer, we set a threshold on its probability from the QA model to filter out some unreliable results.
  - Benefit
    - Widely adaptable to any argument role or event type
    - Judge if some arguments exist
    - Search for arguments in a document (not within a sentence)

RolePred 3: Argument Role Selection

- Role Filtering
  - Judge the salience of an argument role by involving multiple event instances of the same type
    - Ex. *intensity* of the *earthquake* events; *host* for the *award ceremony* events
  - A role name belongs to the event type only if most of the event instances have their associated argument
- Role Merging
  - Different roles can represent similar semantics and share the same arguments in an event
    - Ex. The *date*, *official date*, and *original date* may refer to the same day for a firework event
  - The semantic similarity of two roles is determined by the frequency that they share the same argument in the event instances
    - Ex. Given 10 instances of the firework event, if two roles, *date*, and *official date*, have the same day as their arguments in 5 instances, their similarity is 0.5

Experiment: Argument Role Prediction

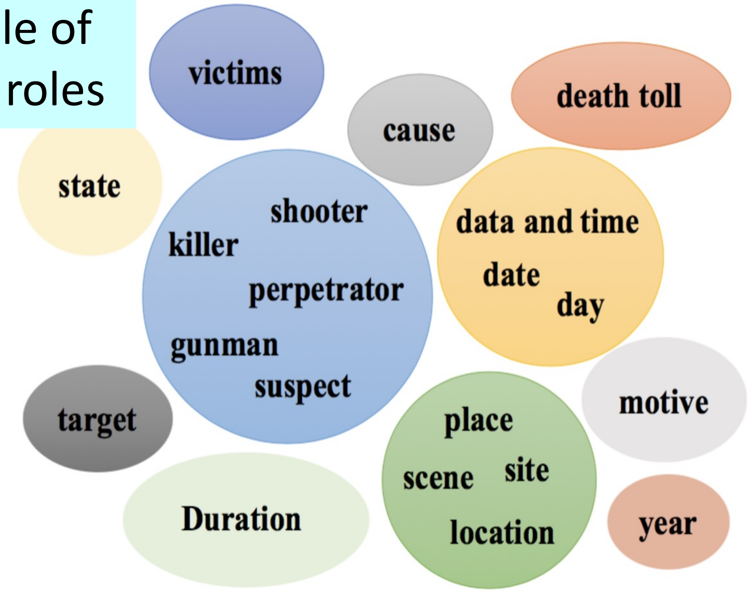
Argument Role Prediction

Models	Hard Matching			Soft Matching		
	Precision	Recall	F1	Precision	Recall	F1
LiberalEE	0.1342	0.2613	0.1773	0.3474	0.5340	0.4209
VASE	0.0926	0.1436	0.1125	0.2581	0.4274	0.3218
ODEE	0.1241	0.3076	0.1768	0.3204	0.4862	0.3862
CLEVE	0.1363	0.2716	0.1815	0.3599	0.5712	0.4415
ROLEPRED (BERT)	0.2128	0.4582	0.2906	0.4188	0.6896	0.5211
ROLEPRED (T5)	**0.2552**	**0.6461**	**0.3659**	**0.4591**	**0.7079**	**0.5570**
- RoleMerge	0.2233	0.6962	0.3381	0.4234	0.7677	0.5457
- RoleMerge - RoleFilter	0.1928	0.6582	0.2983	0.4188	0.7084	0.5264
Human	0.6098	0.8270	0.7020	0.7365	0.8732	0.7990

Argument Extraction w/o Golden Roles

Models	P	R	F1
LiberalEE	0.2009	0.2941	0.2387
VASE	0.2123	0.3257	0.2570
ODEE	0.2402	0.3712	0.2917
CLEVE	0.3529	0.3890	0.3701
ROLEPRED (BERT)	0.4170	0.4333	0.4250
ROLEPRED (Roberta)	**0.4131**	**0.5774**	**0.4817**
- RoleMerge	0.3855	0.6187	0.4750
- RoleMerge - RoleFilter	0.4397	0.5001	0.4679
ROLEPRED (Gold Roles)	0.6664	0.4948	0.5679

An example of generated roles



Extracted events by RolePred and baselines

Output of RolePred

Victims	Maura Binkley and Nancy Van Vessem
State	Florida
Date	November 2, 2018
Killer	Scott Paul Beierle
Place	The yoga studio
Time	5:37 p.m. EDT
Duration	three and a half minutes
Motive	hatred of women
Target	Tallahassee Hot Yoga, a yoga studio
Year	2018

Output of ODEE

Agent	The gunman
Patient	six women

Output of CLEVE

Agent	Scott Paul Beierle
Patient	six women
Time	2018

Outline

- Span Detection
- Entity Typing
- Relation and Event Extraction
  - Relation Extraction
  - Event Discovery
    - Argument Role Prediction
    - Event Chain Mining
      - Unsupervised Event Chain Mining from Multiple Documents [WWW’23]

Multiple Documents Share Salient Events

Multiple News Reports

An intense earthquake hit southern Mexico on Monday, damaging houses, blocking loads and sparking reports of four deaths. "This quake was pretty strong. There are houses destroyed", said Luis Rivera, governor of the San Marcos region, which was also hit by a 7.4 magnitude earthquake ...

A significant 6.9 magnitude earthquake rocked southern Mexico on Monday killing at least three people - including a newborn baby at a hospital - and injuring dozens. The epicenter was just two kilometers from the Mexican town of Madero, and 200 kilometers from Guatemala City. ...

A strong earthquake hit the border of Mexico on Monday, killing at least three people, including a newborn boy, damaging dozens of houses and blocking roads. This quake was pretty strong. Families in the area are really scared because of the whole experience of 2012...

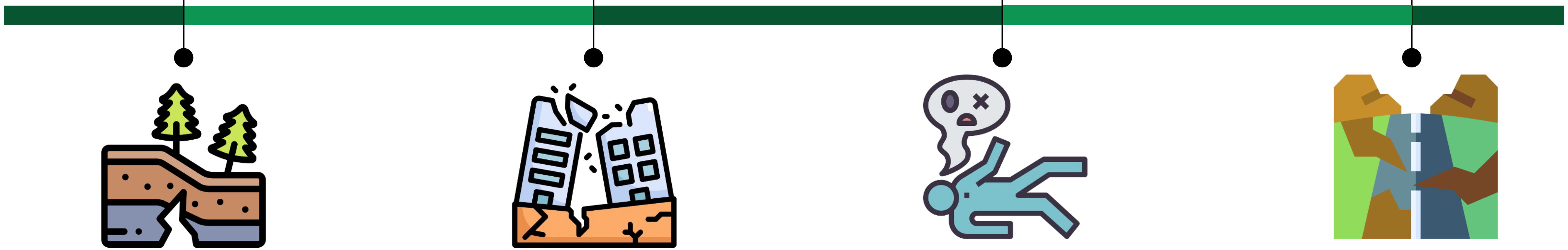
News reports can be summarized as the chains of salient events in temporal order.

Event Chain Mining

Multiple Documents

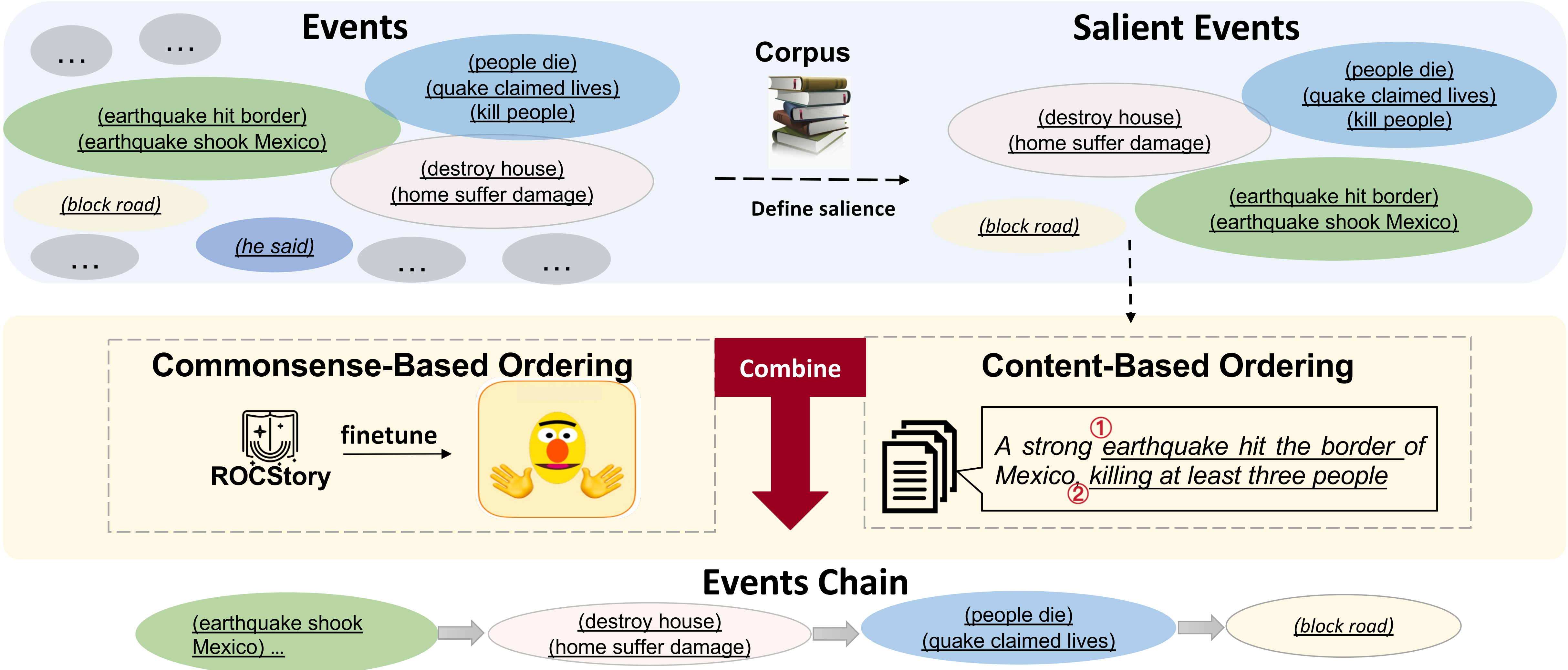
Event Chain

Earthquake hit Mexico | Damage buildings | Kill people | Block roads



Method

- Select salient and informative events
- Arrange salient events in temporal order

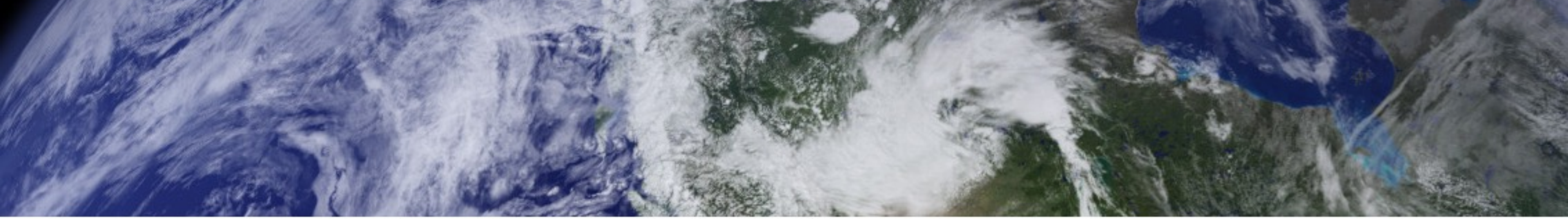


Super Event: Mexico Earthquake in 2017				
Extracted Event Mentions	**Event Cluster**	**Representative Mention in Salient Events**	**Event Chain**	**Reference**
earthquake strike state near coast earthquake felt in city trigger landslide area be fill with vacationer section closed by rockslide they separate to temblor kill people home suffer damage in town …	{earthquake strike state near coast, mexico locate at point, earthquake strike border on monday, evacuate in region, area strike by quake, hit by earthquake, quake strike off coast, quake occur on coast }	trigger landslide kill people family feel scared because experience house destroy block road quake felt in salvador report quake at magnitude suffer disruption to communication	1. hit by earthquake 2. house destroy 3. trigger landslides 4. people died 5. block roads 6. crack open in building	1.earthquake rock Mexico 2.damage houses 3.kill people 4.trigger landslides 5.block roads

Extracted Event Mentions	Event Cluster	Representative Mention in Salient Events	Event Chain	Reference
peterson accused of fires, fire led investigation, he arrested on allegations, peterson entered plea at court, peterson hails from community, scenario rounded with megawatts, organization preserved parcels, transaction culminated years, peterson suspected of fires,	{firefighter arrested, peterson arrested on charges, firefighter arrested on suspicion, peterson arrested, peterson arrested on Tuesday, peterson arrested after capt., he arrested on allegations, } {destroy structures, destroyed by fire, destroying structures, damaging by flames, destroyed by fire, structures claimed by fire, burn homes}	arrest came after months turn into fire body found inside home destroy dozen homes left person injured peterson face years for charge suffer losses fire burn throughout area	1. arrest came after months 2. turn into fire 3. body found inside home 4. destroy dozen homes 5. leave person injured 6. peterson faces years for charge 7. suffer losses 8. fire burn throughout area	1. man started fires 2. fire destroyed dozen homes 3. fire left person dead 4. officers arrested man on suspicion 5. man entered plea at court 6. fire burned miles over week 7. fire fanned by winds

References

- Xiaotao Gu , Zihan Wang , Zhenyu Bi , Yu Meng, Liyuan Liu, Jiawei Han, Jingbo Shang. “UCPhrase: Unsupervised Context-aware Quality Phrase Tagging.” (KDD’21)
- Xiaotao Gu, Yikang Shen, Jiaming Shen, Jingbo Shang, Jiawei Han, “Phrase-aware Unsupervised Constituency Parsing” (ACL’22)
- Jiaxin Huang, Yu Meng, and Jiawei Han, “Few-Shot Fine-Grained Entity Typing with Automatic Label Interpretation and Instance Generation”, (KDD’22)
- Tanay Komarlu, Minhao Jiang, Xuan Wang, Jiawei Han, “ONTOTYPE: Ontology-Guided Annotation-Free Fine-Grained Entity Typing”, (Arxiv’22)
- Yiqing Xie, Jiaming Shen, Sha Li, Yuning Mao, Jiawei Han, “EIDER: Evidence-enhanced Document-level Relation Extraction”, (ACL’22 Findings)
- Sizhe Zhou, Suyu Ge, Jiawei Han, “Corpus-Based Relation Extraction by Identifying and Refining Relation Patterns”, (ECMLPKDD’23)
- Yizhu Jiao, Sha Li, Yiqing Xie, Ming Zhong, Heng Ji and Jiawei Han “Open-Vocabulary Argument Role Prediction for Event Extraction”, (EMNLP’22)
- Yizhu Jiao, Ming Zhong, Jiaming Shen, Yunyi Zhang, Chao Zhang and Jiawei Han, “Unsupervised Event Chain Mining from Multiple Documents”, in Proc. 2023 The Web Conf. (WWW’23)



Q&A