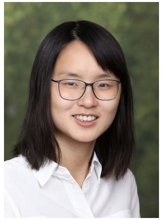
PRINCETON NLP

A Frustratingly Easy Approach for Entity and Relation Extraction

Zexuan Zhong



Danqi Chen

Princeton University

Entity and Relation Extraction: Problem Definition

Input

Bill Smith was in the hotel room

Named Entity Recognition

Bill Smith	hotel	room
PER	FAC	FAC

Relation Extraction

Bill Smith (PER) $\xrightarrow{\text{PHYS}}$ room (FAC)

hotel (FAC) $\xleftarrow{\text{PART-WHOLE}}$ room (FAC)

Entity and Relation Extraction: Problem Definition

Input: a piece of unstructured text

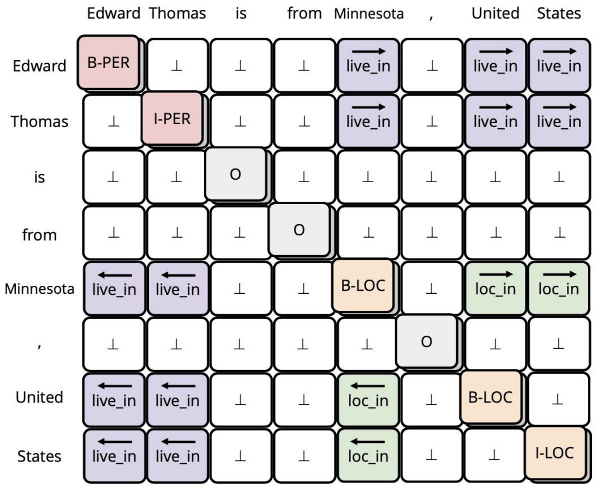
- A sequence of tokens $X = x_1, \ldots, x_n$
  - a set of spans $S = \{s_1, \ldots, s_m\}$

Output:

- A set of entities: $Y_e = \{(s_i, e) : s_i \in S, e \in \mathcal{E}\}$
  - *s*: span, *e*: entity type
- A set of relations: $Y_r = \{(s_i, s_j, r) : s_i, s_j \in S, r \in \mathcal{R}\}$
  - *s*: subject/object span, *r*: relation type

Existing Approaches (2014+)

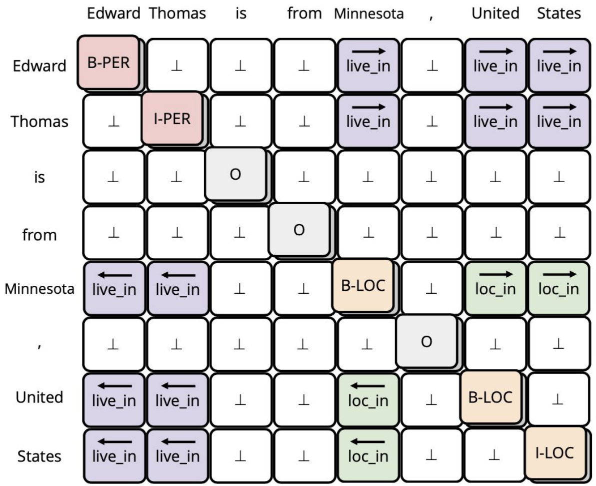
Existing Approaches (2014+)



Structured Prediction

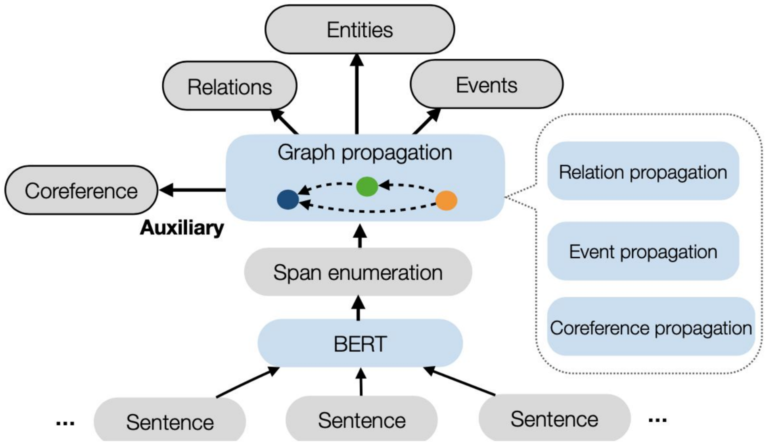
Li and Ji, 2014; Zhang et al., 2017; Katiyar and Cardie, 2017;
Li et al., 2019; Wang and Lu, 2020

Existing Approaches (2014+)



Structured Prediction

Li and Ji, 2014; Zhang et al., 2017; Katiyar and Cardie, 2017; Li et al., 2019; Wang and Lu, 2020



Multi-task Learning

Miwa and Bansal, 2016; Bekoulis et al., 2018; Luan et al., 2019; Wadden et al., 2019; Lin et al., 2020

This Work

1. **Our model: PURE**
    - A **pipelined** approach outperforming all previous joint models!
2. **Why does it work well?**
    - **Understanding** modeling choices between entity and relation extraction
3. **An efficient approximation model w/ large speedup**

This Work

1. **Our model: PURE**
   - A **pipelined** approach outperforming all previous joint models!
2. **Why does it work well?**
   - **Understanding** modeling choices between entity and relation extraction
3. **An efficient approximation model w/ large speedup**

Entity Model

Input

Bill Smith was in the hotel room

Output

Bill Smith
PER

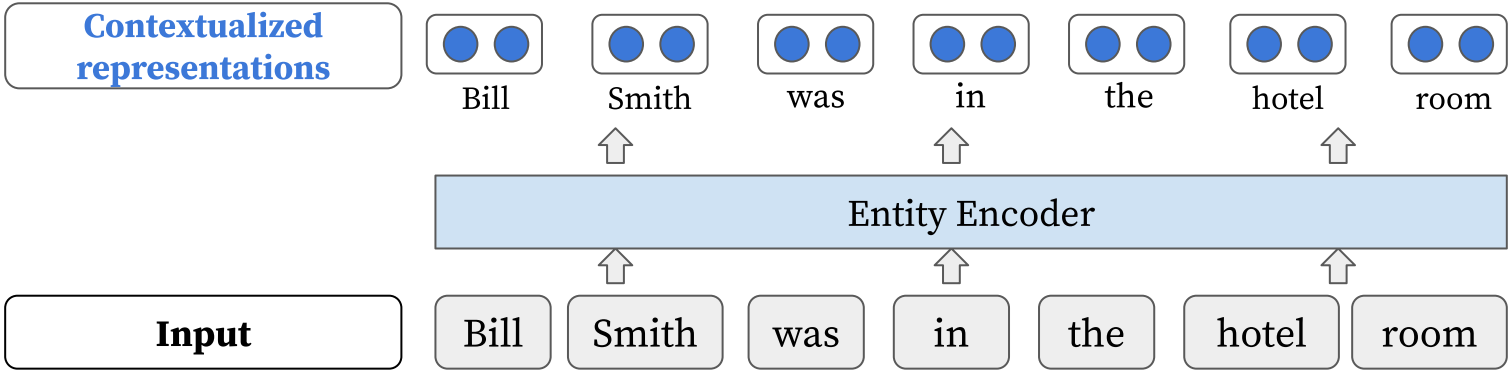
hotel
FAC

room
FAC

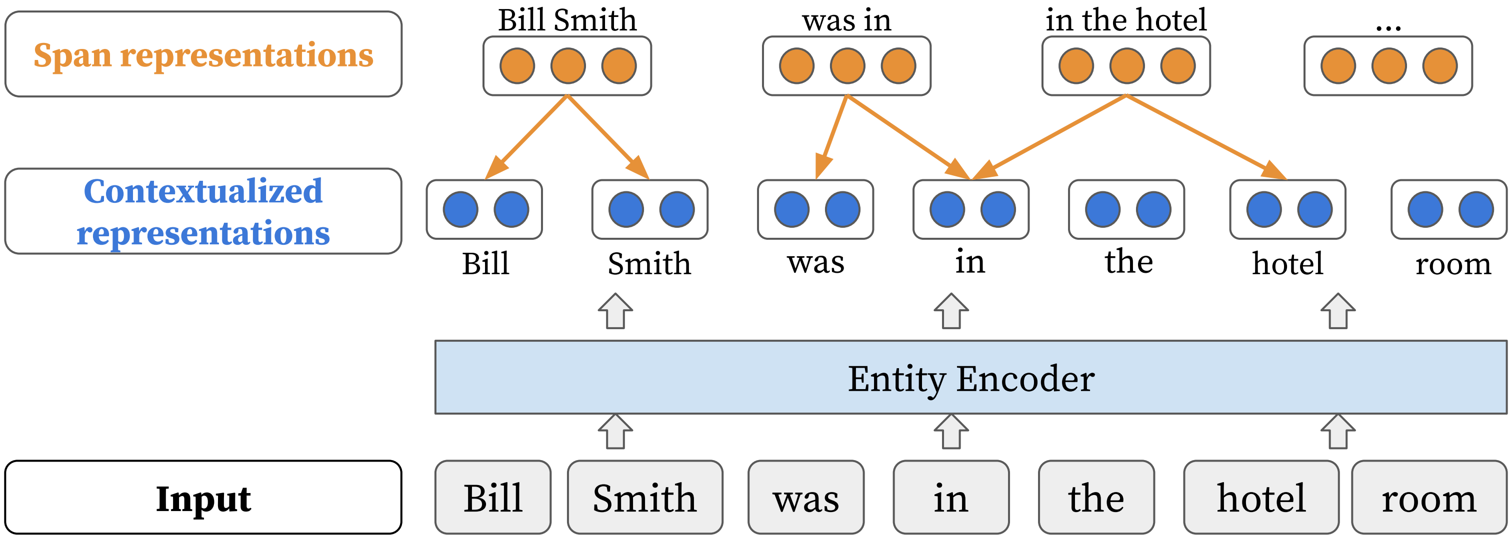
Entity Model

Input	Bill	Smith	was	in	the	hotel	room

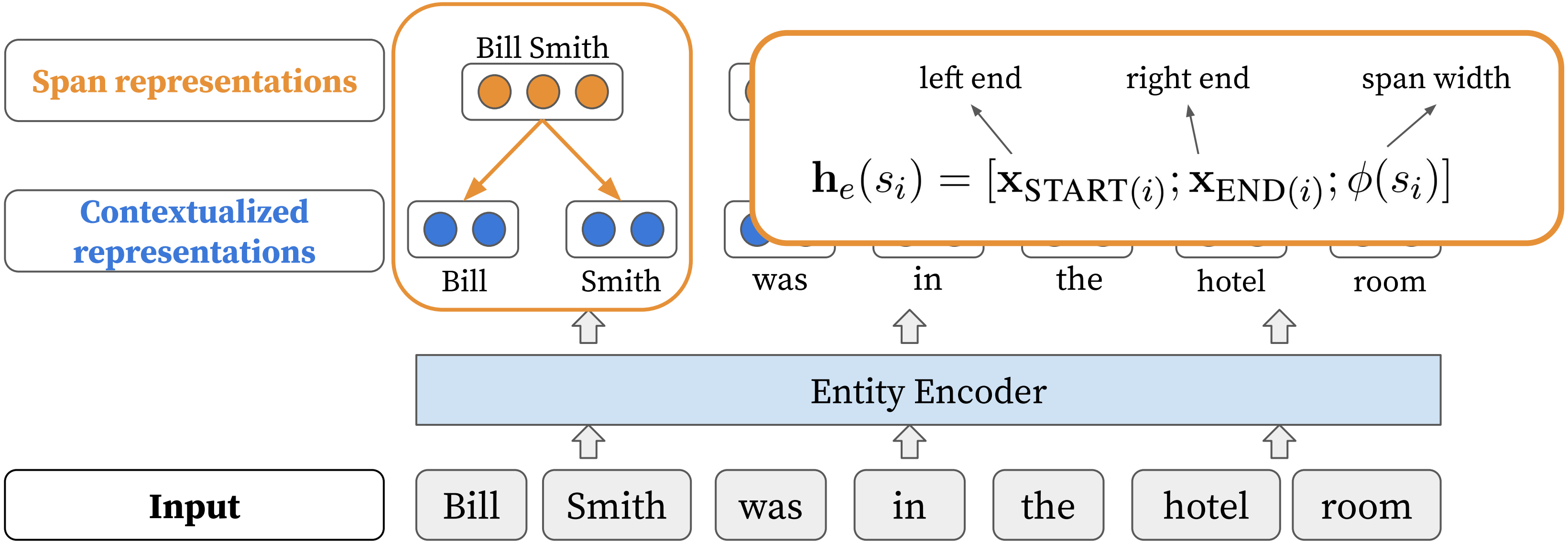
Entity Model



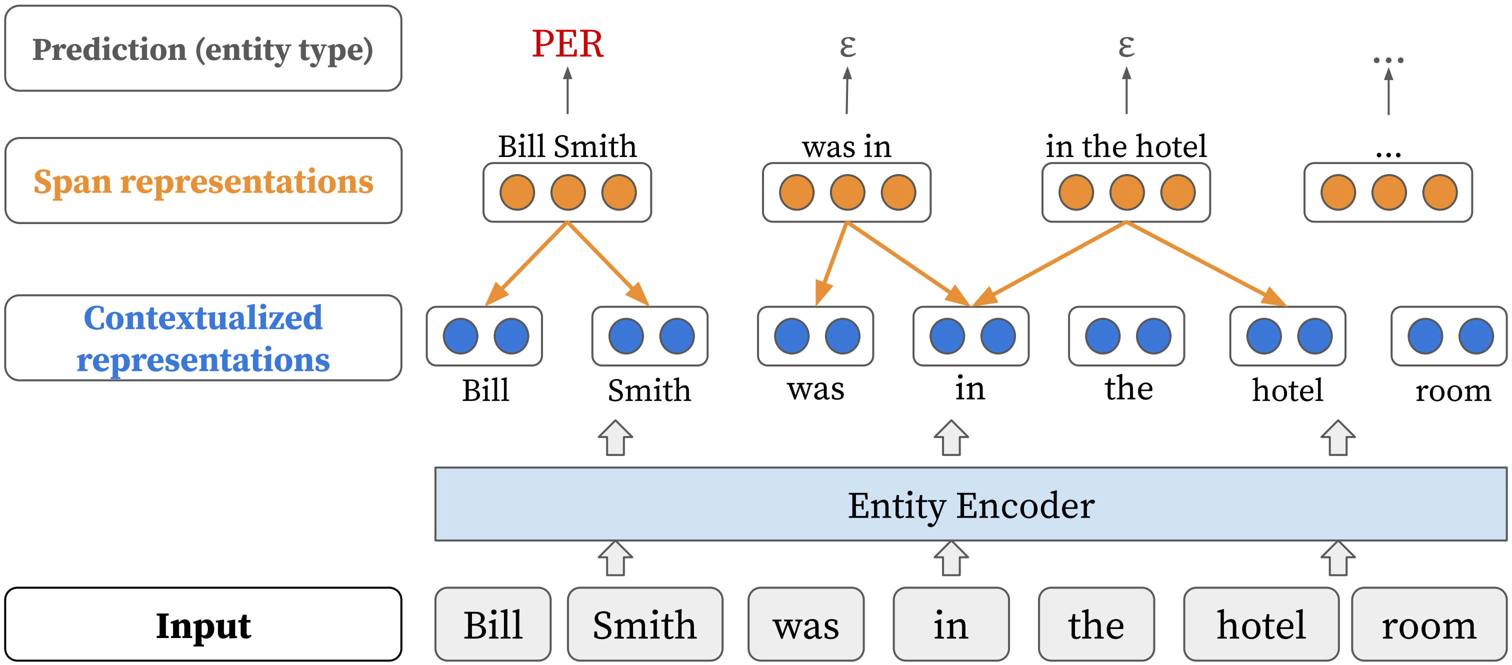
Entity Model



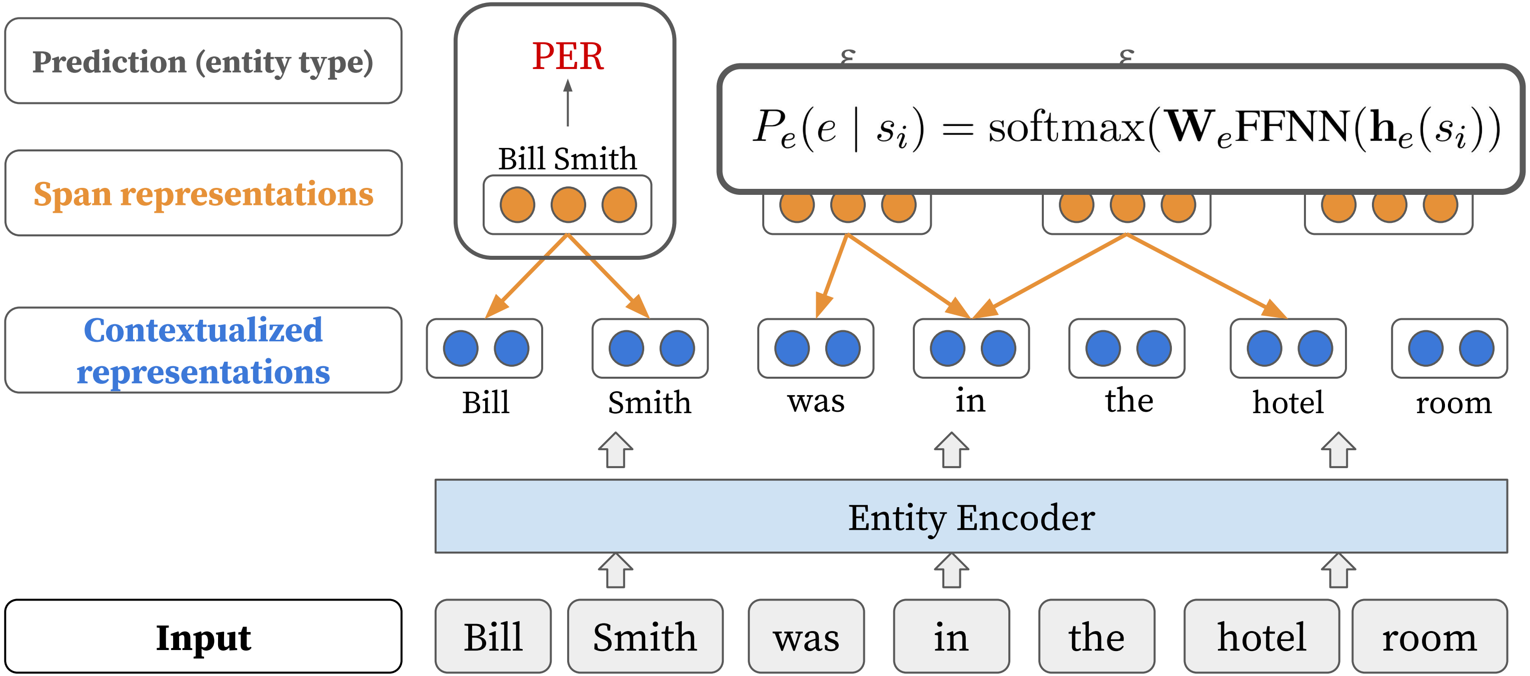
Entity Model



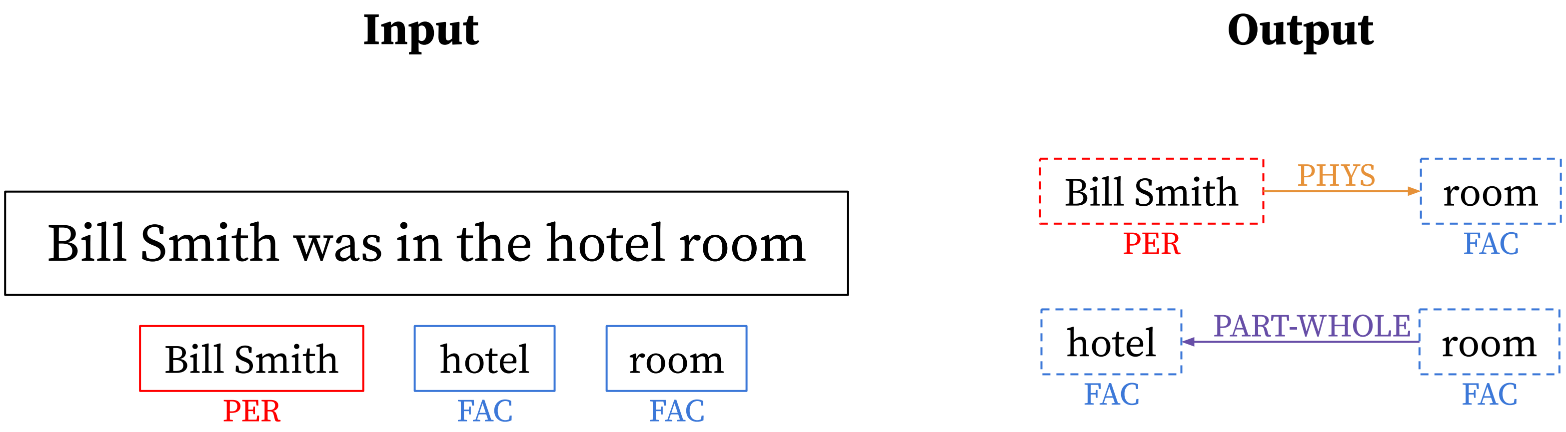
Entity Model



Entity Model



Relation Model



Relation Model: Inserting Markers

Bill Smith was in the hotel room

PER FAC FAC

Bill Smith hotel

PER FAC

Relation Model: Inserting Markers

Bill Smith was in the hotel room

PER FAC FAC

Bill Smith hotel

PER FAC

[S:PER] Bill Smith [/S:PER] was in [O:FAC] hotel [/O:FAC] room

Relation Model: Inserting Markers

Bill Smith was in the hotel room

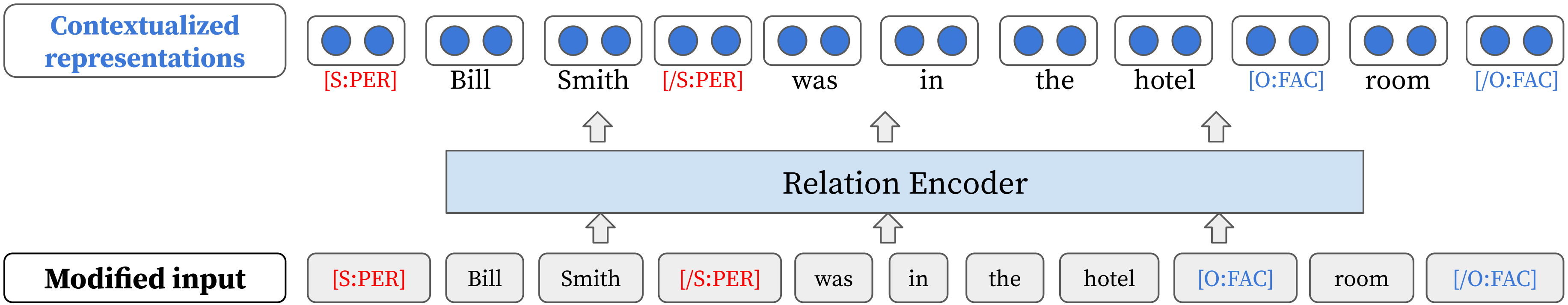
PER FAC FAC

Entity pair		Input with markers
Bill Smith (PER)	hotel (FAC)	[S:PER] Bill Smith [/S:PER] was in [O:FAC] hotel [/O:FAC] room
Bill Smith (PER)	room (FAC)	[S:PER] Bill Smith [/S:PER] was in hotel [O:FAC] room [/O:FAC]
hotel (FAC)	room (FAC)	Bill Smith was in [S:FAC] hotel [/S:FAC] [O:FAC] room [/O:FAC]
…		…

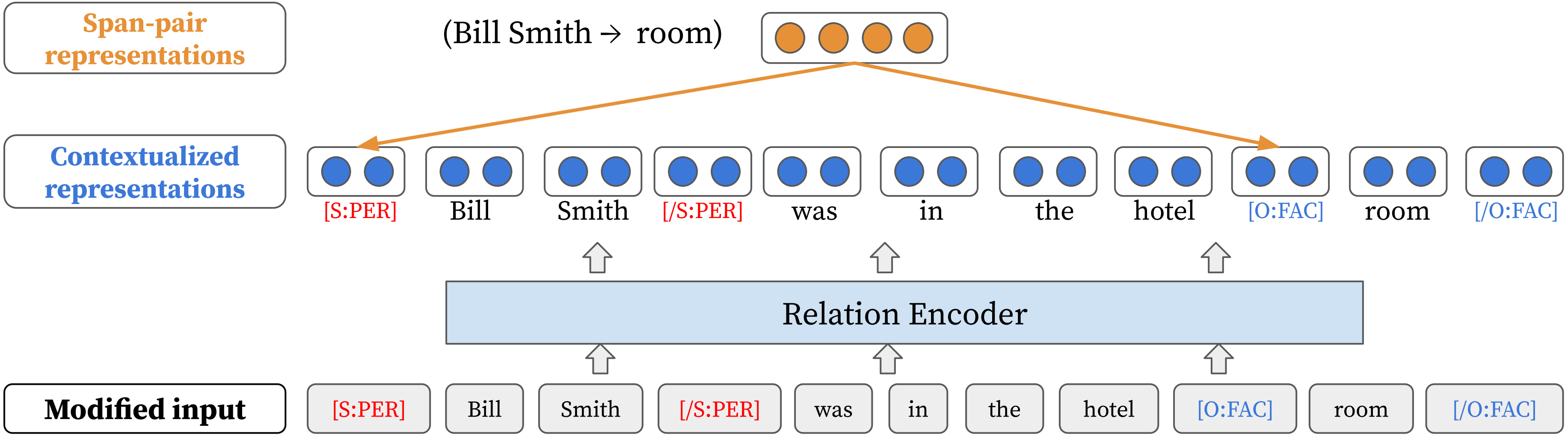
Relation Model

Modified input	[S:PER]	Bill	Smith	[/S:PER]	was	in	the	hotel	[O:FAC]	room	[/O:FAC]

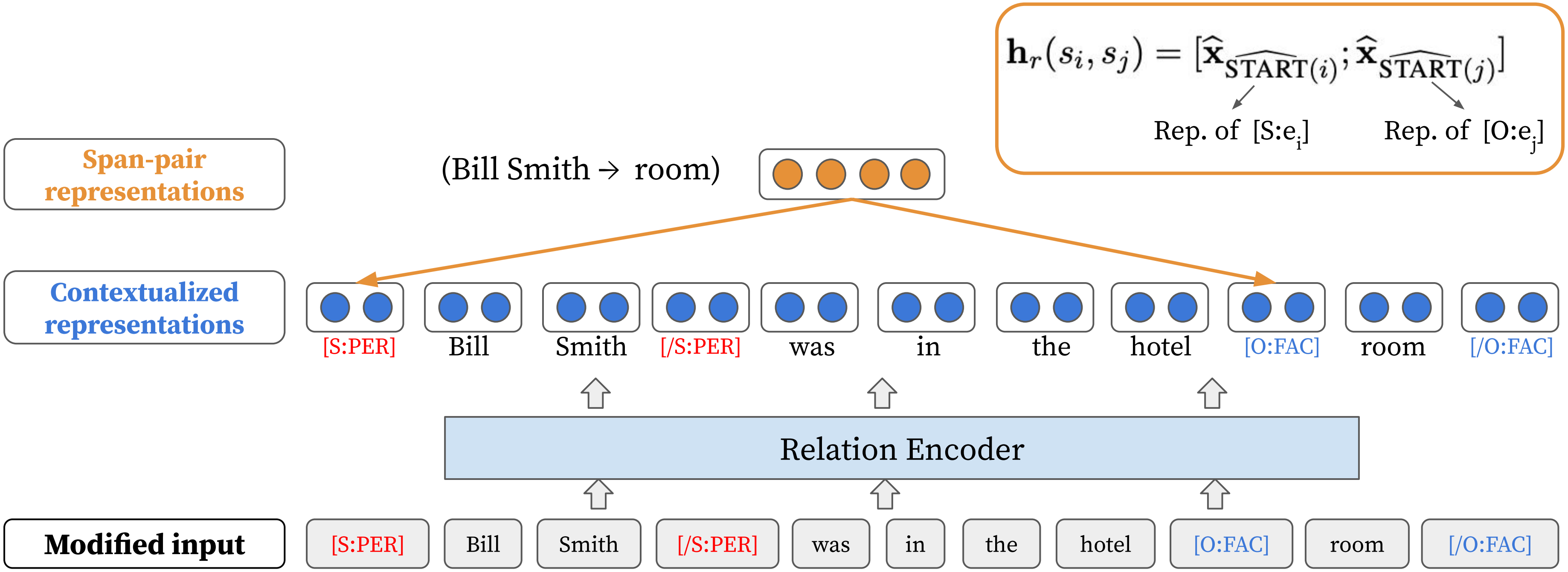
Relation Model



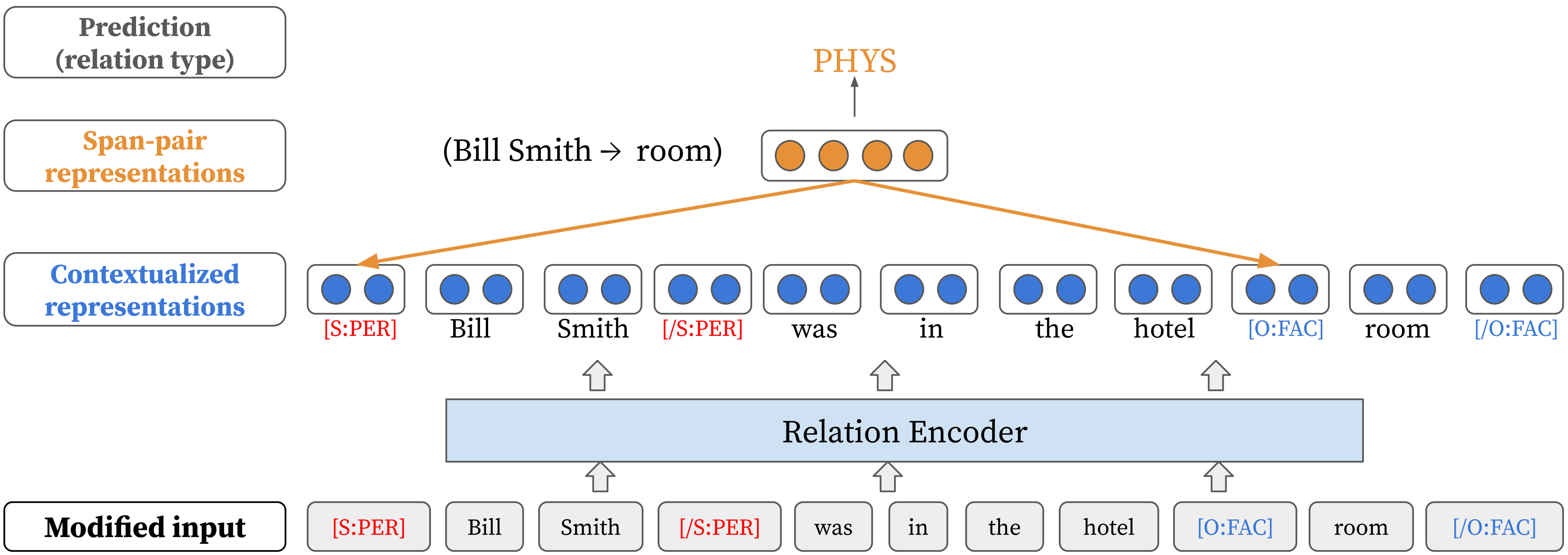
Relation Model



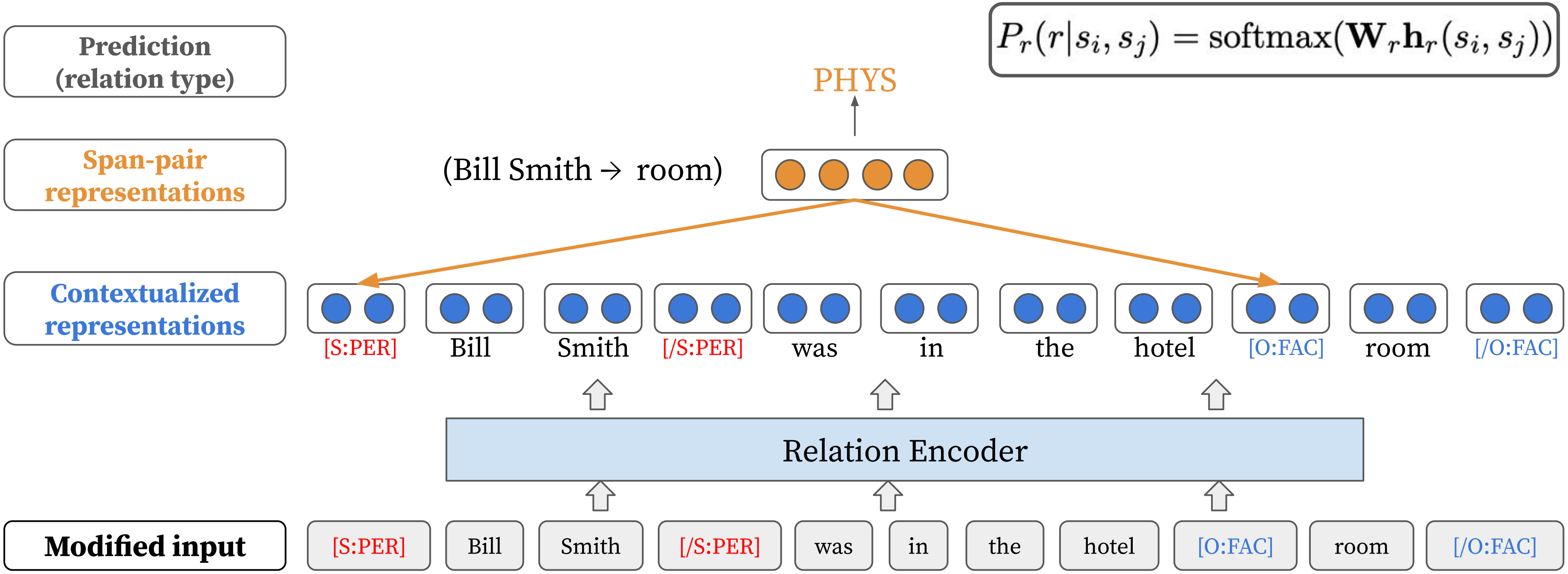
Relation Model



Relation Model



Relation Model



Experimental Settings

Datasets

- ACE04, ACE05: newswire, online forums
- SciERC: scientific articles

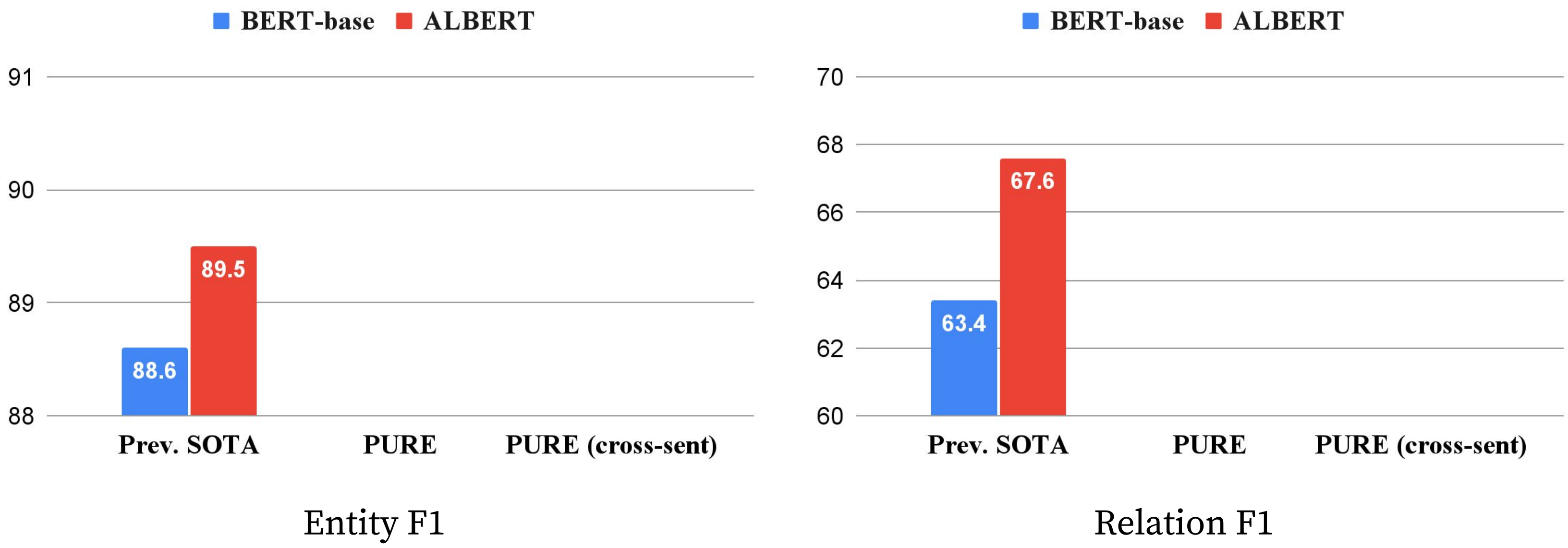
Evaluation metrics

- Entity F1
- Relation F1

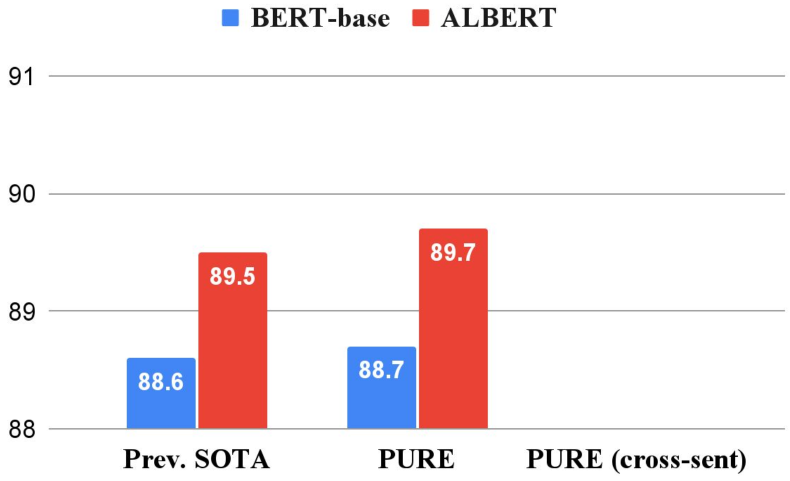
Context information

- Single-sentence
- Cross-sentence

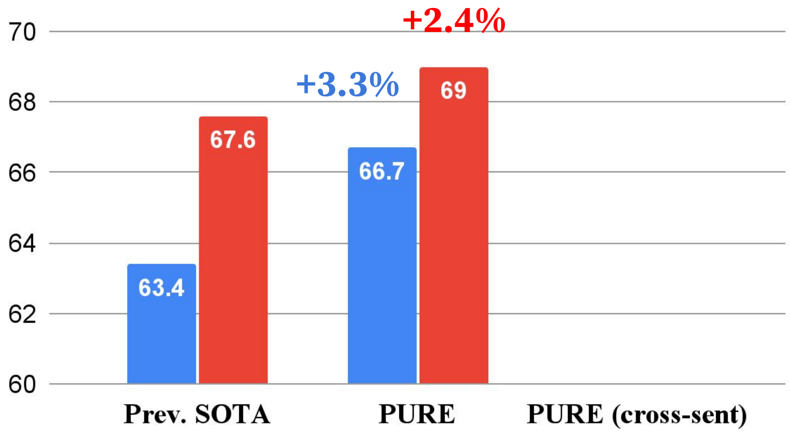
Results on ACE05



Results on ACE05



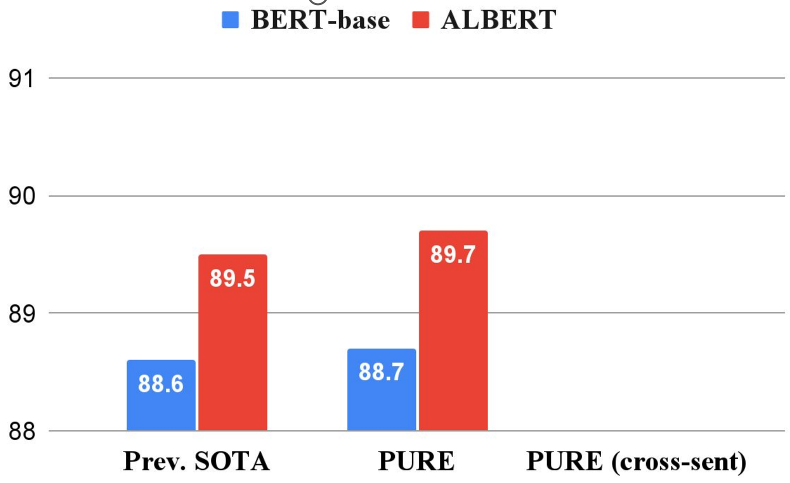
Entity F1



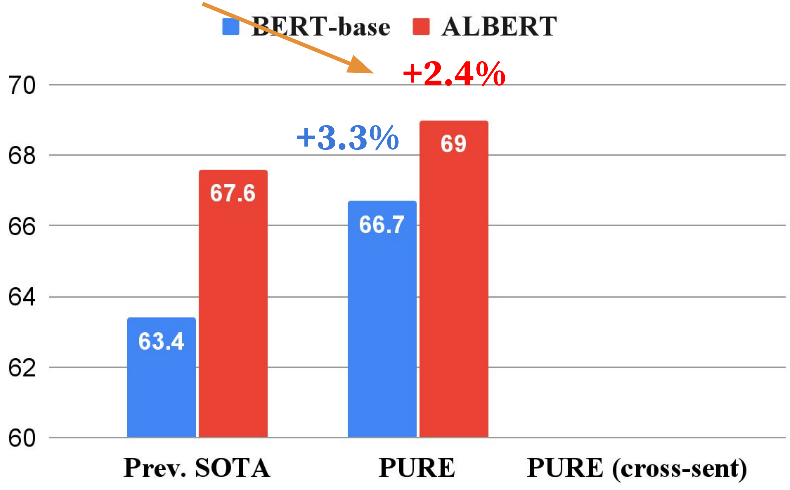
Relation F1

Results on ACE05

With the same pre-trained encoder, PURE significantly outperforms previous SOTA.

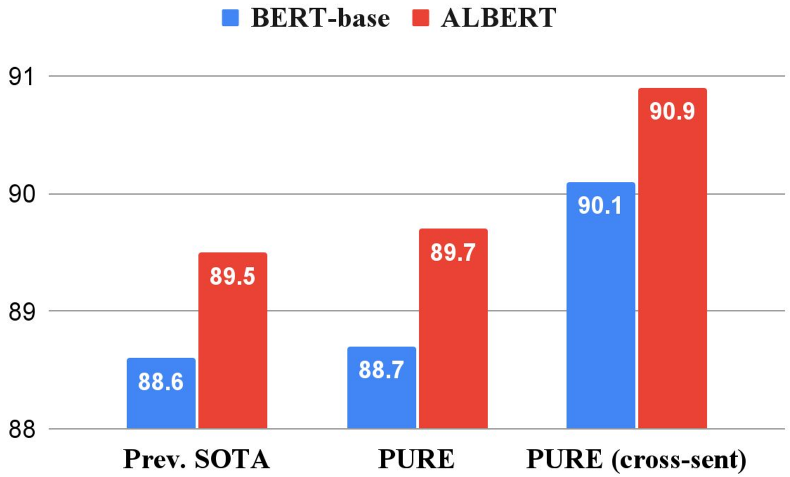


Entity F1

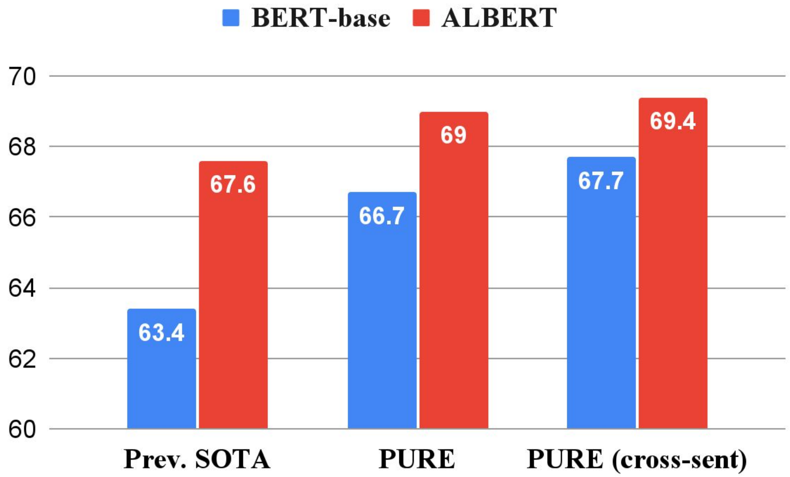


Relation F1

Results on ACE05



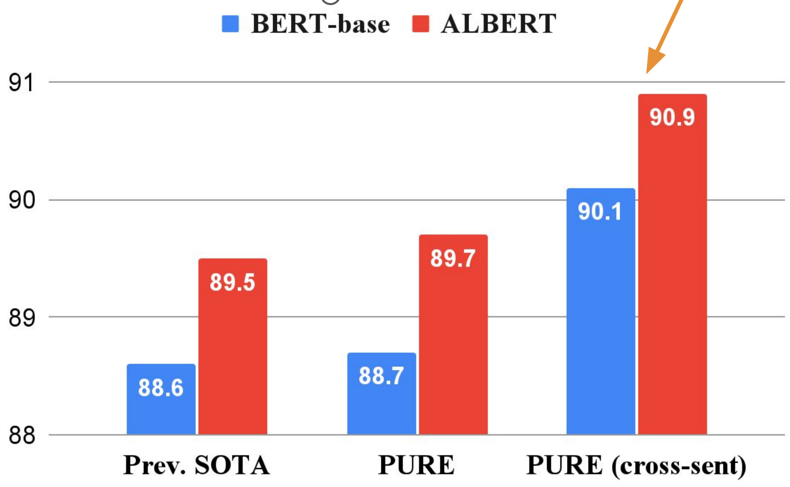
Entity F1



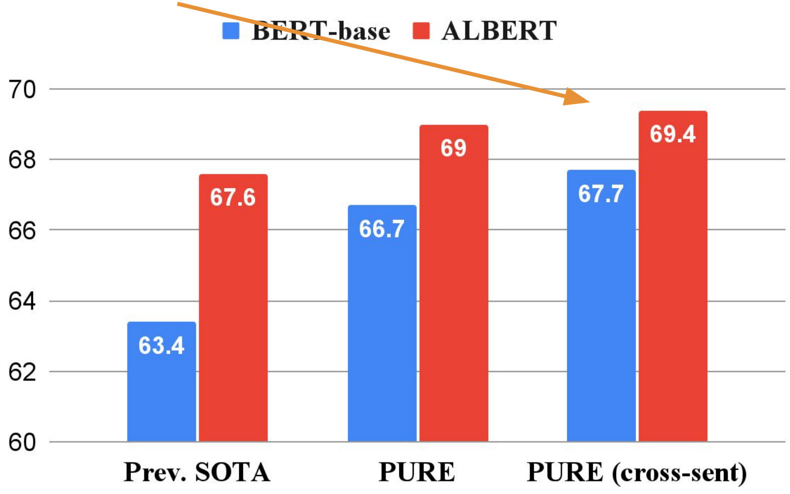
Relation F1

Results on ACE05

Incorporating cross-sentence information can further improve the performance.

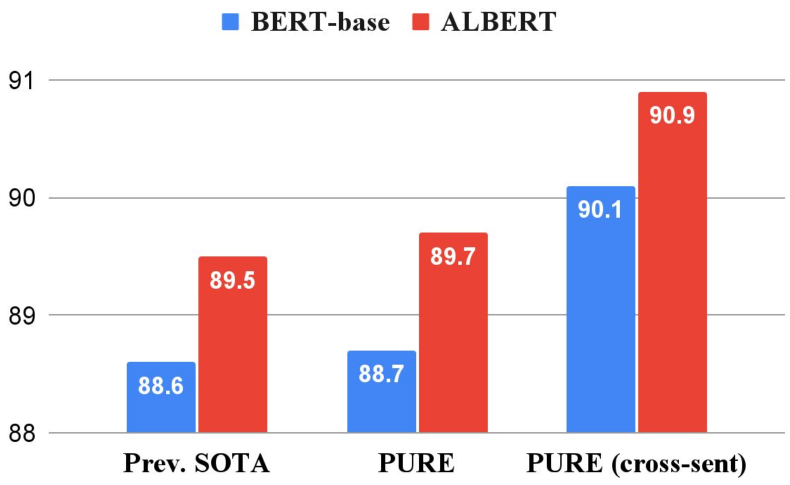


Entity F1

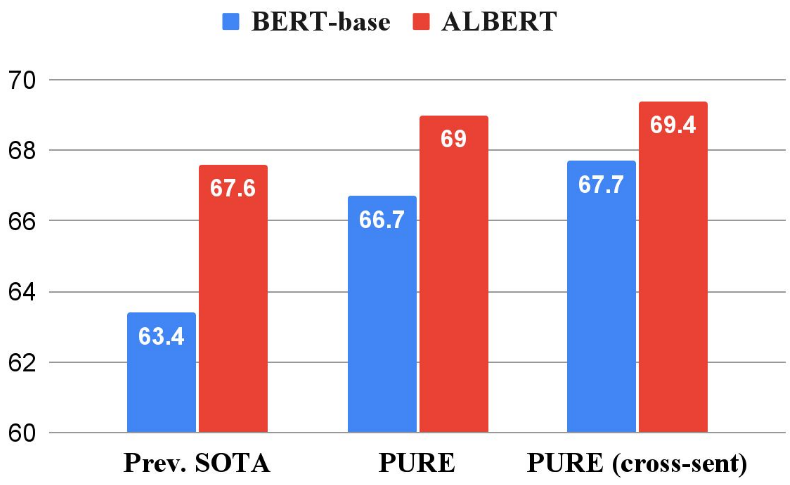


Relation F1

Results on ACE05

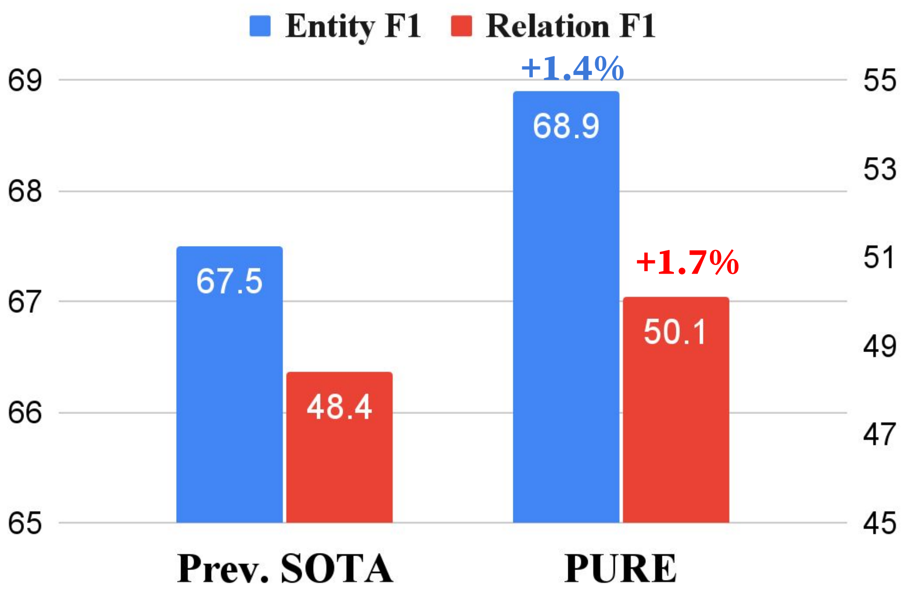


Entity F1



Relation F1

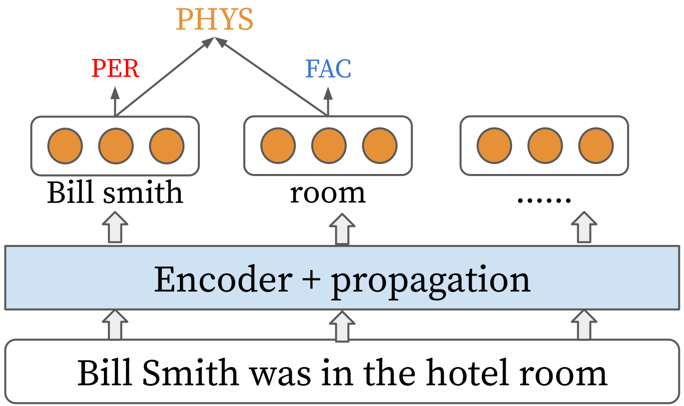
Results on SciERC



This Work

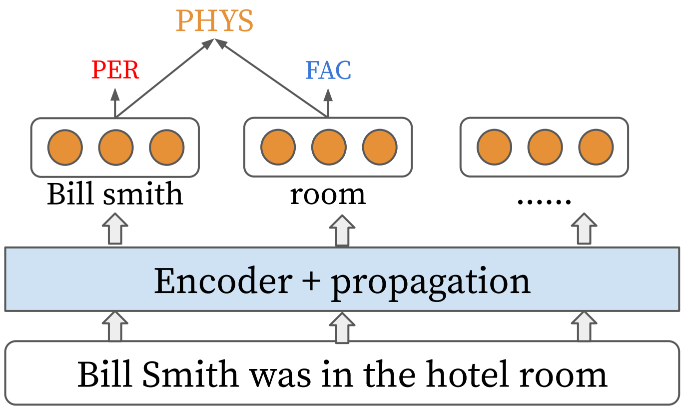
1. **Our model: PURE**
    - A **pipelined** approach outperforming all previous joint models!
2. **Why does it work well?**
    - **Understanding** modeling choices between entity and relation extraction
3. **An efficient approximation model w/ large speedup**

Comparison: Out Approach vs Joint Models

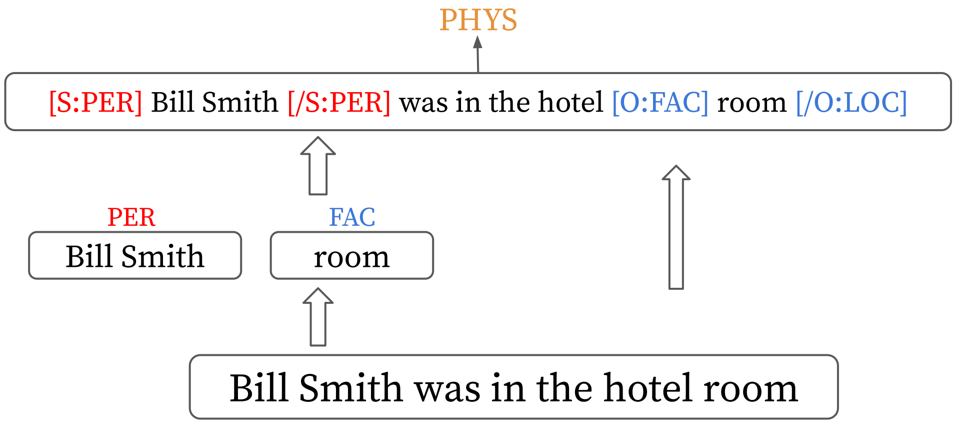


DYGIE++
(Wadden et al. 2019)

Comparison: Out Approach vs Joint Models



DYGIE++
(Wadden et al. 2019)



PURE (ours)

Importance of Typed Markers

Importance of Typed Markers

No marker

PHYS

Bill Smith was in the hotel room

Typed markers

PHYS

[S:PER] Bill Smith [/S:PER] was in the hotel [O:FAC] room [/O:FAC]

Importance of Typed Markers

No marker

PHYS

Bill Smith was in the hotel room

Untyped markers

PHYS

[S] Bill Smith [/S] was in the hotel [O] room [/O]

Typed markers

PHYS

[S:PER] Bill Smith [/S:PER] was in the hotel [O:FAC] room [/O:FAC]

Importance of Typed Markers

No marker

PHYS

Bill Smith was in the hotel room

Untyped markers

PHYS

[S] Bill Smith [/S] was in the hotel [O] room [/O]

Markers + entity auxiliary loss

PER PHYS FAC

[S] Bill Smith [/S] was in the hotel [O] room [/O]

Typed markers

PHYS

[S:PER] Bill Smith [/S:PER] was in the hotel [O:FAC] room [/O:FAC]

Importance of Typed Markers

Relation F1

No marker 67.6%

PHYS

Bill Smith was in the hotel room

Untyped markers 70.5%

PHYS

[S] Bill Smith [/S] was in the hotel [O] room [/O]

Markers + entity auxiliary loss 70.7%

PER PHYS FAC

[S] Bill Smith [/S] was in the hotel [O] room [/O]

Typed markers **72.6%**

PHYS

[S:PER] Bill Smith [/S:PER] was in the hotel [O:FAC] room [/O:FAC]

Why Pipelined Model?

Why Pipelined Model?

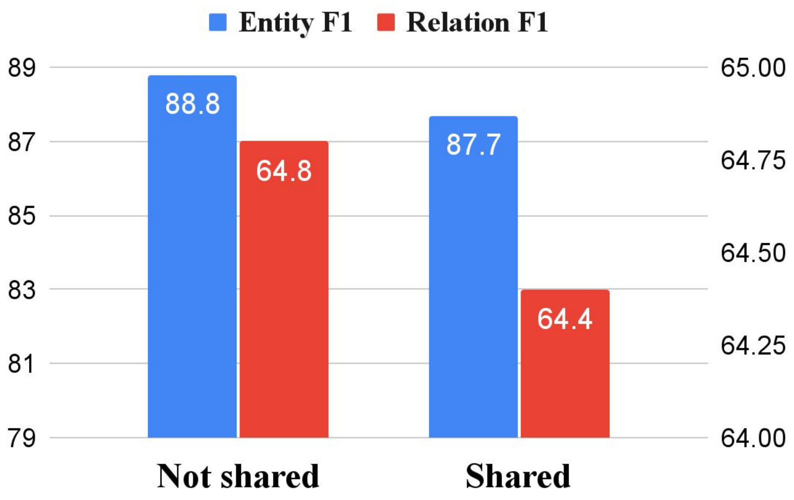
1. Two encoders capture distinct information!

    Does **sharing encoders** help?

Why Pipelined Model?

1. Two encoders capture distinct information!

Does **sharing encoders** help? **No!**



Why Pipelined Model?

2. Modeling entity-relation interaction

Why Pipelined Model?

2. Modeling entity-relation interaction

Entity → Relation? 🙂

- Use typed markers!

Why Pipelined Model?

2. Modeling entity-relation interaction

Entity → Relation? 🙂

- Use typed markers!

Relation → Entity? 🙁

- Adding a relation auxiliary in the entity model does not help!

Why Pipelined Model?

3. Error propagation?

Why Pipelined Model?

3. Error propagation?

- Both **jackknifing** and **beam pruning** (Lee et al., 2017; Luan et al., 2018) didn't improve performance!

This Work

1. **Our model: PURE**
    - A **pipelined** approach outperforming all previous joint models!
2. **Why does it work well?**
    - **Understanding** modeling choices between entity and relation extraction
3. **An efficient approximation model w/ large speedup**

Improving Runtime Efficiency of PURE

Bill Smith was in the hotel room

PER FAC FAC

M predicted entities

Improving Runtime Efficiency of PURE

Bill Smith was in the hotel room

PER FAC FAC

M predicted entities

$O(M^2)$ entity pairs!

[S:PER] Bill Smith [/S:PER] was in [O:FAC] hotel [/O:FAC] room

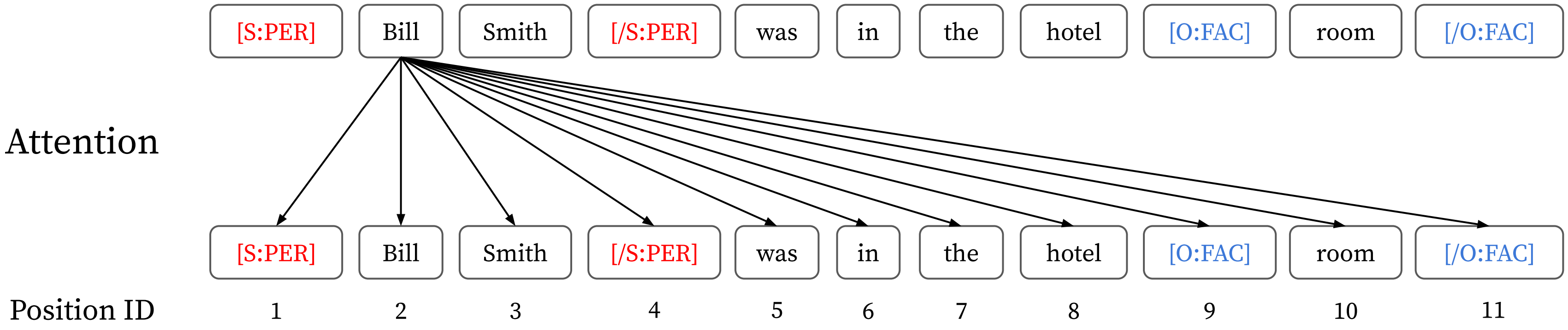
[S:PER] Bill Smith [/S:PER] was in hotel [O:FAC] room [/O:FAC]

Bill Smith was in [S:FAC] hotel [/S:FAC] [O:FAC] room [/O:FAC]

...

Addressing the Efficiency Issue

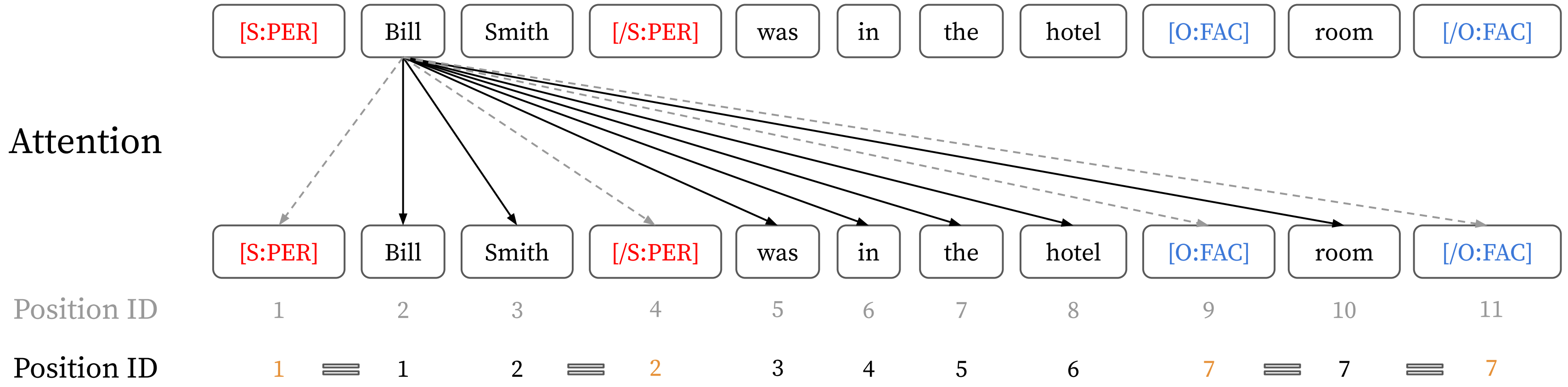
Key idea: re-use computations for different pairs of spans in the same sentence



Original Model

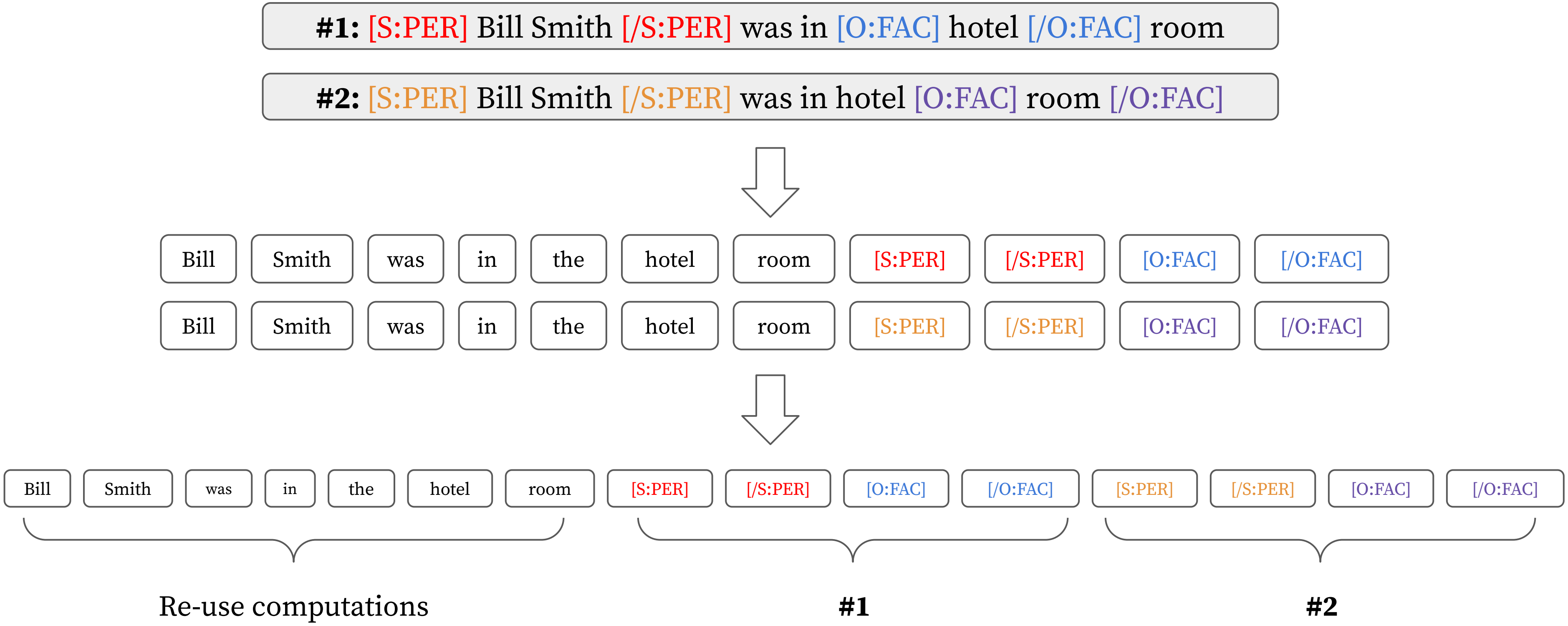
Addressing the Efficiency Issue

Key idea: re-use computations for different pairs of spans in the same sentence



Approximation Model

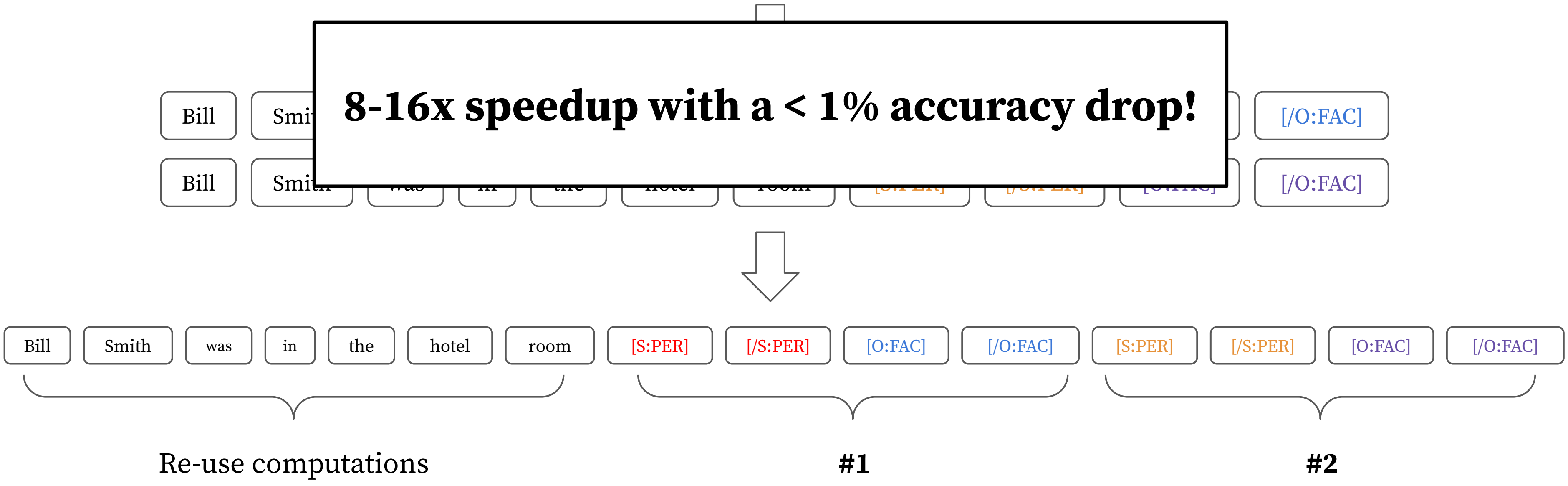
Approximation Model with Batch Computations



Approximation Model with Batch Computations

#1: [S:PER] Bill Smith [/S:PER] was in [O:FAC] hotel [/O:FAC] room

#2: [S:PER] Bill Smith [/S:PER] was in hotel [O:FAC] room [/O:FAC]



Conclusions

PURE: A simple and effective approach for entity and relation extraction

- Learns two independent encoders
- Outperforms all previous joint model on three datasets
- An efficient approximation: 8-16x speedup with a small accuracy drop

Conclusions

PURE: A simple and effective approach for entity and relation extraction

- Learns two independent encoders
- Outperforms all previous joint model on three datasets
- An efficient approximation: 8-16x speedup with a small accuracy drop

Let's rethink the value of joint training in entity and relation extraction!

PRINCETON NLP

Thank You!

Paper: https://arxiv.org/pdf/2010.12812.pdf

Code: https://github.com/princeton-nlp/PURE