Implicit Argument Prediction with Event Knowledge

Pengxiang Cheng Katrin Erk
pxcheng@cs.utexas.edu katrin.erk@mail.utexas.edu
The University of Texas at Austin

TEXAS
The University of Texas at Austin

Introduction

Text: More than 2,600 people have been infected by ***Ebola*** in Liberia, Guinea, Sierra Leone and Nigeria since the ***outbreak*** began in December, according to the World Health Organization. Nearly 1,500 have ***died***.
Question: The X ***outbreak*** has ***killed*** nearly 1,500.

- ***Ebola*** is an implicit argument of both ***outbreak*** and ***die***, which is key to answering this question.
- Implicit arguments are NOT syntactically connected to their predicates, thus hard to extract.
- Previous work focused on very small datasets [1].

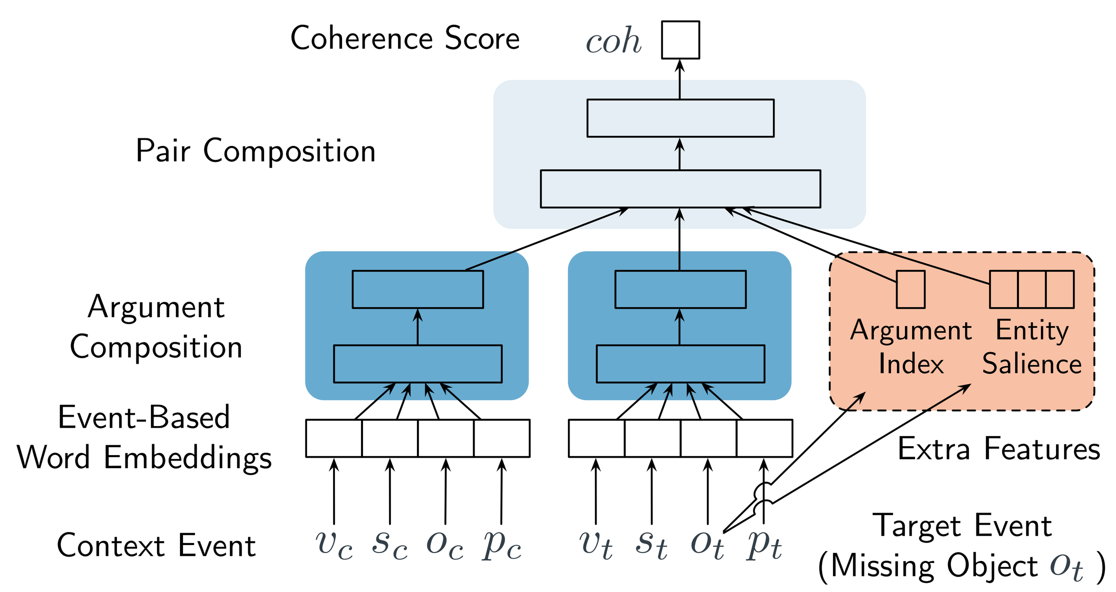
Methods

- **Event knowledge** is key to implicit argument detection: We select candidate x_j with the highest narrative coherence score S_j:

$$S_j = \max_{c=1,\ldots,n\ c\neq t} coh\left(e_t(j), e_c\right), \quad j = 1, \ldots, m$$

where e_c are context events, $e_t(j)$ is the target event with candidate x_j filling in as the implicit argument.

- We compute the coherence scores between event pairs using a variant of the event composition model [2].



- Implicit arguments tend to be salient entities, so we include **entity salience** features [3].
  - Numbers of named, nominal, pronominal, and total mentions of the entity.

Argument Cloze Task

- We address the data issue by a simple cloze task, for which data can be generated automatically at scale for both training and evaluation.

> Manville Corp. said it will build a $ 24 million power plant to provide electricity to its Igaras pulp and paper mill in Brazil .
>
> The company said the plant will ensure that it has adequate energy for the mill and will reduce the mill's energy costs .

(a) A piece of raw text from OntoNotes corpus.

> x_0 = The company x_1 = mill x_2 = power plant
>
> e_0: (*build-pred*, x_0*-subj*, x_2*-dobj*, —)
> e_1: (*provide-pred*, —, *electricity-dobj*, x_1*-prep_to*)
> e_2: (*ensure-pred*, x_2*-subj*, —, —)
> e_3: (*has-pred*, x_0*-subj*, *energy-dobj*, x_1*-prep_for*)
> e_4: (*reduce-pred*, x_2*-subj*, *cost-dobj*, —)

(b) Extracted events ($e_0 \sim e_4$) and entities ($x_0 \sim x_2$), using gold annotations from OntoNotes.

> e_0, e_2, e_3, e_4: same as above
> e_1: (*provide-pred*, —, *electricity-dobj*, *??-prep_to*)
>
> x_0 = The company x_1 = mill x_2 = power plant

(c) Example of an argument cloze task for *prep_to* of e_1.

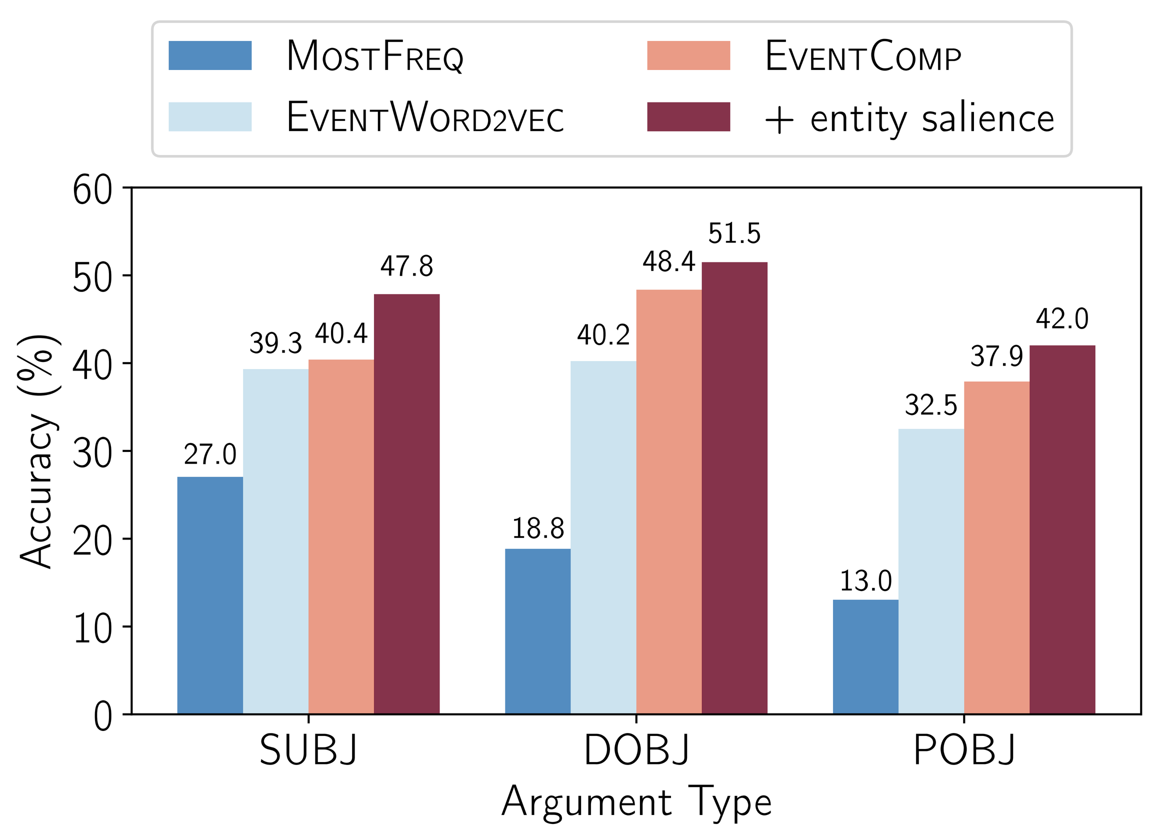
- **Training**: English Wikipedia.
- **Evaluation**: OntoNotes.
  - OntoNotes contains human-labeled dependency and coreference annotation, providing **gold** test data.
  - We construct two datasets, ON-SHORT and ON-LONG.

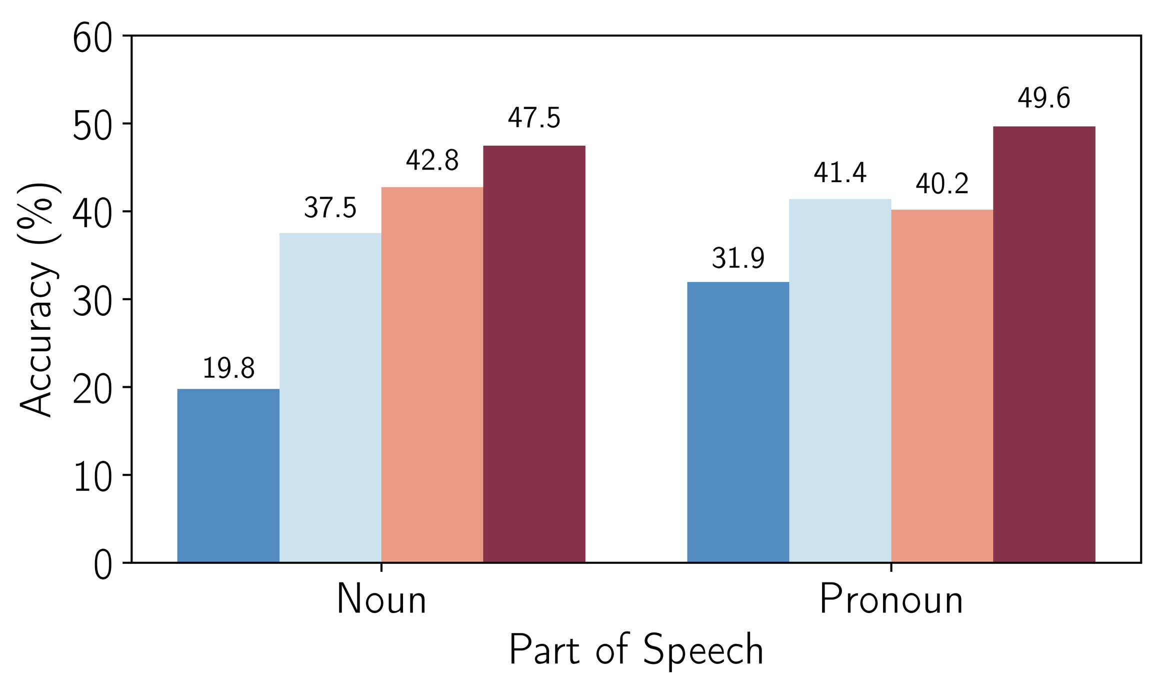
Results on OntoNotes

- We compare our model with 3 baselines.
- ON-LONG is significantly harder than ON-SHORT, with much longer documents and much more candidates.

Accuracy (%)	ON-SHORT	ON-LONG
RANDOM	8.29	2.71
MOSTFREQ	22.76	17.23
EVENTWORD2VEC	38.40	21.49
EVENTCOMP	41.89	21.79
+ entity salience	**47.75**	**27.87**

- We break down the results by argument type and part of speech of the implicit argument.





Results on G&C

- G&C [1] is a human-labeled implicit argument dataset.
- Less than 1,000 examples on 10 nominal predicates.

	PP	R	F_1
Gerber & Chai (2012) [1]	57.9	44.5	50.3
GCAUTO	49.9	40.1	44.5
EVENTCOMP	46.7	47.3	47.0
+ entity salience	49.3	49.9	**49.6**

Conclusion

- Neural model with event knowledge has superior performance on both synthetic and natural data.
- Entity salience is important throughout for performance.

References

[1] Matthew Gerber and Joyce Y. Chai. "Semantic Role Labeling of Implicit Arguments for Nominal Predicates". In: *Computational Linguistics* 38.4 (2012).
[2] Mark Granroth-Wilding and Stephen Clark. "What Happens Next? Event Prediction Using a Compositional Neural Network Model". In: *AAAI*. 2016.
[3] Jesse Dunietz and Daniel Gillick. "A New Entity Salience Task with Millions of Training Examples". In: *EACL*. 2014.

Acknowledgements

This research was supported by NSF grant IIS 1523637. We also acknowledge the Texas Advanced Computing Center for providing grid resources that contributed to these results, and we would like to thank the anonymous reviewers for their valuable feedback.

- Code available at `https://github.com/pxch/event_imp_arg`.