A sequence-to-sequence approach for document-level relation extraction

John Giorgi, Gary D. Bader, Bo Wang

October 16, 2022

Introduction

- Novel end-to-end joint learning approach for inter-sentence relation extraction.[1]
- Utilizes sequence to sequence architecture.
- Representation schema for coreferent entities and *n*-ary relations.

[1] Document-level is a stretch, due to encoder limit of 512 tokens they did paragraphs.

Introduction

- New benchmarks for end-to-end results over some biomedical datasets.
- Competitive results against more complex architectures for datasets with established end-to-end results.

Defining Terms

End-to-end RE:

- Relation extraction depends on entities.
- Pipeline methods (current standard), use one or more models for NER, and one or more models for RE over discovered entities.
- End-to-end approaches use one model (possibly with a classification head) to discover the relations, relying on internal representations to jointly extract and implicitly coordinate entity and relation information.

NB: The authors use *pipeline* to refer to the RE component. In NER/RE practice, pipeline usually refers to the whole system, NER component included.

Defining Terms

Coreference:

- The same entity may have one or more mentions in a given text unit (type vs. token).
- If a relation holds between two entites, how to reflect this for each entity's mentions?

Defining Terms

Sequence to sequence (seq2seq):

- Encoder to decoder.
- Encoder maps each input token to a contextual representation.
- Decoder maps each encoder token output and prior context to an output token.
- Sequence cross-entropy loss used in training.

Motivation

- Lots of entity and relation information at the document and cross document level.
- Generalizing sentential pipeline methods (the current standard) for inter-sentential RE is involved.[2]
- Lots of information takes the form of *n*-ary relations, tricky to reconstruct this from binary relations.

[2] e.g. our NER/RE system for radiotherapy.

Datasets

- **CDR**
  Chemical-induced disease (CID) relations, binary relations.
- **GDA**
  Gene-disease associations, binary relations.
- **DGM**
  Drug-gene-mutations, ternary relations.
- **DocRED** General domain, binary relations.

Datasets

Table 6: Evaluation datasets used in this paper with details about their annotations. Inter-sentence relations (%) are the fraction of relations in the test set that cross sentence boundaries. We consider a relation intra-sentence if any sentence in the document contains at least one mention of each entity in the relation, and inter-sentence otherwise. *This differs from the estimate in Yao et al. (2019), see Appendix B.

Corpus	Nested Mentions?	Discontinuous Mentions?	Coreferent mentions?	n-ary relations?	Inter-sentence relations (%)
CDR (Li et al., 2016b)	✓	✓	✓	✗	29.8
GDA (Wu et al., 2019)	✓	✗	✓	✗	15.6
DGM (Jia et al., 2019)	✗	✗	✓	✓	63.5
DocRED (Yao et al., 2019)	✗	✗	✓	✗	12.5*

Linearization Schema

X: Variants in the **estrogen receptor alpha (ESR1)** gene and its mRNA contribute to risk for **schizophrenia**.

Y: `estrogen receptor alpha ; ESR1 @GENE@ schizophrenia @DISEASE@ @GDA@`

Full schema:

$< entity\ mention_{1,1} >$; ... ; $< entity\ mention_{1,n} >$ @ $< entity\ type_1 > \ldots$ $< entity\ mention_{m,1} >$; ... ; $< entity\ mention_{m,k} >$ @ $< entity\ type_m >$ @ $< relation\ type >$

Model Structure

- Seq2seq architectre.
- Decoder: Single-layer LSTM with randomly initialized weights.
- Encoder: PubMedBERT on DGM, GDA, and CDR. $\text{BERT}_{\text{BASE}}$ for DocRED.
- Decoder generates special @START and @END tokens, along with entity and relation types.
- Decoder includes entity tokens in ouput via a including okens from the input text in output vocabulary[3].
- 6 head cross attention mechanism[4] between encoder and decoder.

[3] *copy mechanism* in the authors' terms.

[4] https://vaclavkosar.com/ml/cross-attention-in-transformer-architecture

RE on Gold Entities with Entity Hinting

X: estrogen receptor alpha ; ESR1 @GENE@ schizophrenia @DISEASE@ @SEP@ Variants in the estrogen receptor alpha (ESR1) gene and its mRNA contribute to risk for schizophrenia.

Full schema:
$< \textit{entity mention}_{1,1} >$; ... ; $< \textit{entity mention}_{1,n} >$ @ $< \textit{entity type}_1 > \dots$
$< \textit{entity mention}_{m,1} >$; ... ; $< \textit{entity mention}_{m,k} >$ @ $< \textit{entity type}_m > \dots$
@SEP $< \textit{input text} >$

n-ary Relations (DGM)

Method	P	R	F1
Jia et al. (2019) †	62.9	76.2	68.9
seq2rel (entity hinting)	**84.0**	**84.8**	**84.4**
seq2rel (entity hinting, relaxed)	84.1	84.9	84.5
seq2rel (end-to-end)	68.9	65.9	67.4
seq2rel (end-to-end, relaxed)	78.3	74.9	76.6

DGM has ternary relations. Jia et al. (2019) uses multiscale architecture (uses multiple representations over different sizes of text spans and types of sub-relations). Both use gold entities (entity hinting in seq2rel case).

RE with Gold Entities (CDR, GDA)

	CDR			GDA		
Method	P	R	F1	P	R	F1
Christopoulou et al. (2019)	62.1	65.2	63.6	–	–	81.5
Nan et al. (2020)	–	–	64.8	–	–	82.2
Minh Tran et al. (2020)	–	–	66.1	–	–	82.8
Lai and Lu (2021)	64.9	67.1	66.0	–	–	–
Xu et al. (2021)	–	–	68.7	–	–	83.7
Zhou et al. (2021)	–	–	**69.4**	–	–	83.9
seq2rel (entity hinting)	68.2	66.2	67.2	84.4	85.3	**84.9**
seq2rel (entity hinting, relaxed)	68.2	66.2	67.2	84.5	85.4	85.0
seq2rel (end-to-end)	43.5	37.5	40.2	55.0	55.4	55.2
seq2rel (end-to-end, relaxed)	56.6	48.8	52.4	70.3	70.8	70.5

(Not enough room, full breakdown in paper appendix)

End-to-end RE (DocRED)

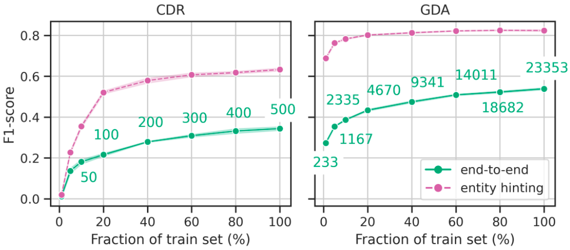
Method	P	R	F1
JEREX (Eberts and Ulges, 2021)	42.8	**38.2**	**40.4**
seq2rel (end-to-end)	**44.0**	33.8	38.2
seq2rel (end-to-end, relaxed)	53.7	41.3	46.7

Eberts and Ulges (2021) use JEREX. Extends BERT with four task-specific components that use BERTs outputs to per- form entity mention localization, coreference resolution, entity classification, and relation classification. They present two versions of their relation classifier, denoted "global re- lation classifier" (GRC) and "multi-instance relation classifier" (MRC). The authors compare against JEREX-MRC in DocRED end to end.

End-to End RE Ablation (CDR, DocRED)

	CDR				DocRED			
	P	R	F1	Δ	P	R	F1	Δ
seq2rel (end-to-end)	**41.0**	35.1	37.8	–	46.9	**36.1**	**40.8**	–
- pretraining	9.4	6.9	8.0	-29.8	18.5	7.7	10.8	-30.0
- fine-tuning	24.3	20.5	22.2	-15.6	42.4	15.5	22.7	-18.1
- vocab restriction	39.6	32.2	35.5	-2.3	45.2	35.5	39.7	-1.1
- sorting relations	36.1	29.2	32.3	-5.6	**52.9**	17.4	26.2	-14.7
+ constrained decoding	40.8	**35.6**	**38.0**	+0.2	46.8	35.9	40.6	-0.2

Training Set Size vs. Performance (CDR, GDA)



Conclusion

See paper for bibliography.