Joint Entity and Relation Extraction from Scientific Documents: Role of Linguistic Information and Entity Types

TYSS Santosh (Microsoft), Prantika Chakraborty (IACS), Sudakshina Dutta (IIT Goa),

Debarshi Kumar Sanyal (IACS), Partha Pratim Das (IIT Kharagpur)

2nd Workshop on Extraction and Evaluation of Knowledge Entities from Scientific Documents
EEKE2021 @ JCDL2021

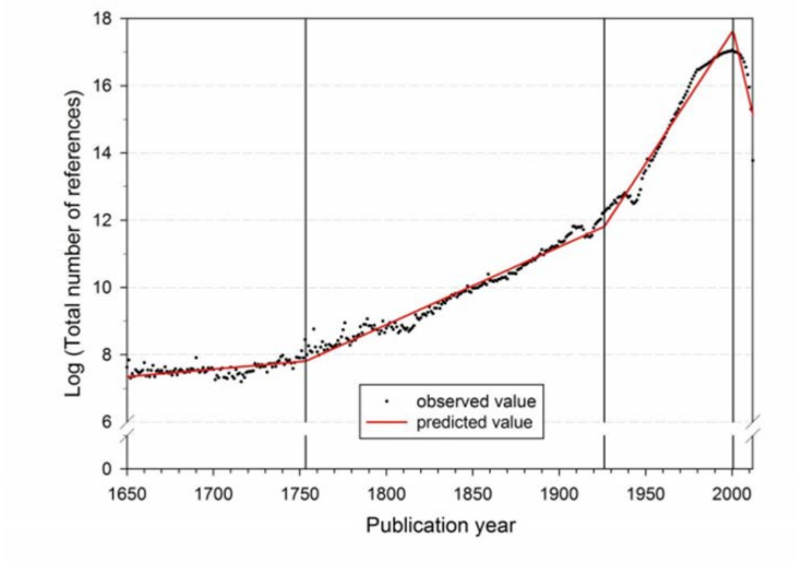
Entity & Relation in Scientific Documents

- 'The generalized LR parsing$_{\text{Method}}$ is enhanced in this approach$_{\text{Generic}}$.'
- Relation(generalized LR parsing, approach) = Used-for.

- Named entity recognition (NER): Extraction of entities, i.e., entity *mentions* and their *types*.
- Relation extraction (RE): Extraction of the relation between entity pairs.

Why NER and RE?

- Scientific output roughly doubles every 9 years.
- Automated reading & analysis of papers is the need of the hour.
- Requires NER and RE.
- Useful in knowledge graph construction, entity search, question answering, etc.



Segmented growth of the annual number of cited references from 1650 to 2012 (citing publications from 1980 to 2012)

Nature News Blog, 2014.

Literature Review

- Research on NER and RE has been around for many years.
- Focus on NER and RE from scientific documents is relatively recent.
- Notable initiative:
  - Semeval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers. [15]

"... The purpose of the task is to automatically identify relevant domain-specific semantic relations in a corpus of scientific publications. ..."

Literature Review

- 2 approaches for NER & RE:
  - Pipelined: NER followed by RE. Simple but does not use cross-dependency.
  - Joint: NER and RE formulated as a joint task with a global optimization objective. Uses inter-connection between tasks. But can propagate error.
- Most NER methods [2,3,12,13,25,30,31] label tokens with BIO/BILOU tags.
- Our method is inspired by SpERT [1] which labels spans instead of tokens. Hence, it supports overlapping entities. SpERT uses a pretrained transformer (BERT) and shallow classifiers on top of it.

Problem Definition

- Input: Sentence D containing tokens $\{d_1, d_2, \ldots d_n\}$.
- Let s_i denote a span of tokens from D, and S denote all possible spans in D.

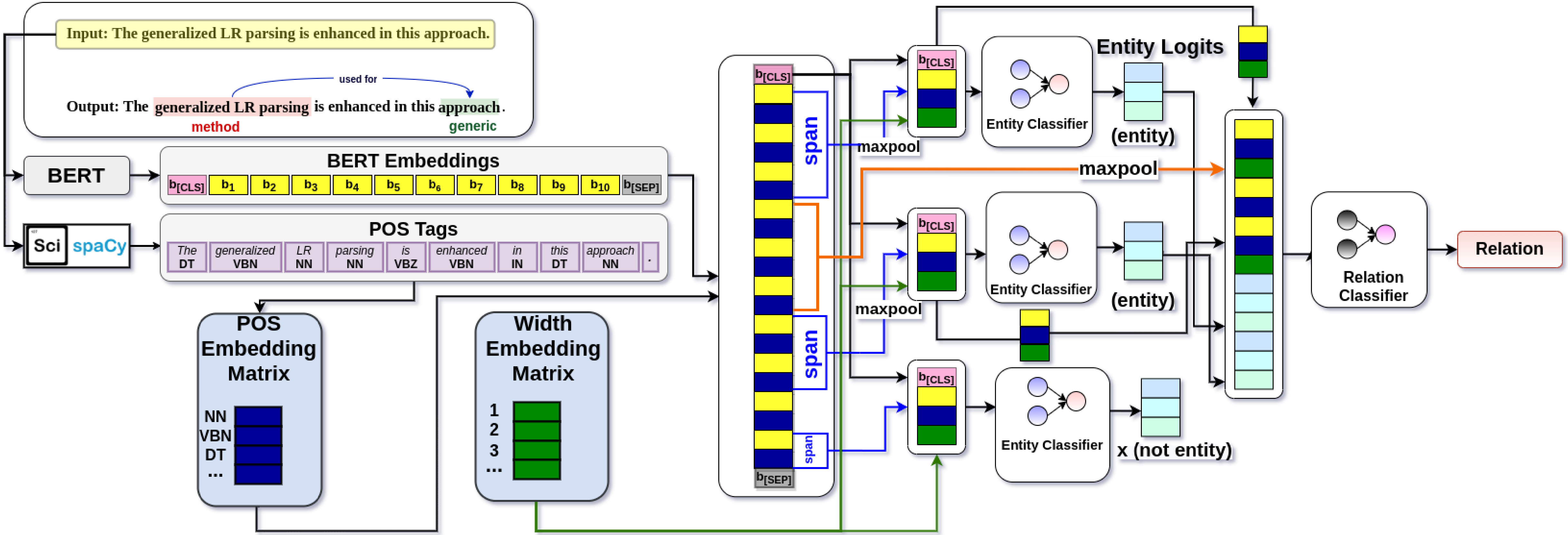
- Let ξ denote a set of pre-defined entity types.
- NER: For each span $s_i \in S$, predict an entity type $e(s_i) \in \xi \cup \{\emptyset\}$ where $\{\emptyset\}$ denotes s_i is not an entity.

- Let λ denote a set of pre-defined relation types.
- RE: For each pair of spans $s_i \in S, s_j \in S$, predict a relation type $r(s_i, s_j) \in \lambda \cup \{\emptyset\}$ where $\{\emptyset\}$ denotes s_i and s_j do not share a relation.

Proposed Approach

1. Pretrained Transformer (BERT): Generates embeddings of the input sentence and its constituent tokens.
2. POS Encoder: ScispaCy generates POS tags of tokens. POS embedding matrix converts them embeddings.
3. Fusion Module: Concatenates BERT embeddings of tokens with their POS embeddings.
4. Entity Classifier: Entity representations are formed from every sequence of $k \leq 10$ consecutive tokens and classified by a shallow feed-forward neural network (FFNN), which is the entity classifier.
5. Relation Classifier: Spans *not* predicted as entities are discarded. For every pair of remaining spans, representations of relations are formed, incorporating embeddings of the predicted entity logits, and classified by a shallow FFNN. Relations can be asymmetric.

Proposed Approach



Code: https://github.com/dksanyal/SpERT.PL

Dataset: SciERC

- 500 abstracts of AI papers.
- 6 scientific *entity types*: Task, Method, Metric, Material, Other-Scientific-Term, and Generic.
- 7 *relations*: Compare, Conjunction, Evaluate-For, Used-For, Feature-Of, Part-Of, and Hyponym-Of
- #Sentences: 2,687.
- Training subset: 1,861 + 275 = 2,136 sentences
- Test subset: 551 sentences

Dataset: ADE

- 2 *entity types*: Adverse-Effect and Drug.
- 1 *relation*: Adverse-Effect.
- #Sentences: 4,272; from medical reports.
- #Relations: 6,821 relations.
- Used 10-fold cross validation.
- 2 cases:
    - *With overlap:* all entities and relations are retained.
    - *Without overlap*: around 120 relations with overlapping entities (e.g., \`lithium' is a drug included in \`lithium intoxication') are removed.

Implementation Details

- Pretrained transformer for SciERC: SciBERT (scibert_scivocab_cased)[2]
- Pretrained transformer for ADE: SciBERT (scibert_scivocab_cased), BioBERT (biobert-base-cased-v1.1)[3]
- Dimension of POS embedding = Dimension of span width embedding = 25
- Trained for 20 epochs with Adam optimizer.
- Sigmoid activation threshold in relation classifier = 0.4
- #Negative samples = 100 per sentence.
- Training batch size = 10.

Evaluation Metrics

- NER: An *entity* is considered correct if the entity type and span are predicted correctly.
- RE: Given two text spans, the model also performs RE. Correctness defined in two ways:
  - Strict RE: Relation type and the two related entities (*both span and entity type*) must be correct.
  - Boundaries RE: Relation type and only the spans of the two related entities must be correct.
- Report micro-average for SciERC, both micro- and macro-average for ADE, and only strict RE for ADE. (Since only one relation occurs in ADE, the averaging method for RE does not matter.)

Results: SciERC

Model	NER			Boundaries RE			Strict RE		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
SpERT.PL (SciBERT)	69.82 (±0.44)	**71.25** (±0.51)	**70.53** (±0.37)	51.94 (±0.72)	**50.62** (±0.94)	**51.25** (±0.55)	39.94 (±0.9)	**38.98** (±0.89)	**39.41** (±0.77)
SpERT [1]	**70.87**	69.79	70.33	**53.4**	48.54	50.84	**40.51**	36.82	38.57
DyGIE++ [4]	-	-	67.5	-	-	48.4	-	-	-
DyGIE [5]	-	-	65.2	-	-	41.6	-	-	-
SciIE [6]	67.2	61.5	64.2	47.6	33.5	39.3	-	-	-
PURE (Single sentence) [13]	-	-	66.6	-	-	48.2	-	-	35.6
PURE (Cross sentence) [13]	-	-	68.9	-	-	50.1	-	-	36.8

Performance on SciERC. Micro-average scores are reported.

Results: ADE

	Model	NER (Micro-average)			NER (Macro-average)			Strict RE		
		Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
With Overlap	**SpERT.PL** (BioBERT)	**90.05**	91.69	**90.86**	**90.33**	**92.03**	**91.17**	**80.11**	**84.18**	**82.03**
	SpERT.PL (SciBERT)	89.15	**91.73**	90.4	89.43	91.96	90.72	78.54	83.98	81.16
	SpERT [1]	88.69	89.2	88.95	88.99	89.59	89.28	77.77	79.96	78.84
Without Overlap	**SpERT.PL** (BioBERT)	**90.44**	**91.3**	**90.86**	**90.66**	**91.64**	**91.14**	**80.33**	**84.57**	**82.39**
	SpERT.PL (SciBERT)	89.89	91.16	90.52	89.15	90.75	89.94	79.04	84.39	81.62
	CMAN [12]	-	-	-	-	-	89.4	-	-	81.14
	Table Sequence [11]	-	-	-	-	-	89.7	-	-	80.1
	SpERT [1]	89.02	88.87	88.94	89.26	89.26	89.25	78.09	80.43	79.24
	Relation-Metric [14]*	86.16	88.08	87.1	-	-	-	77.36	77.25	77.29
	Multi-head + AT [7]	-	-	86.7	-	-	-	-	-	75.52
	Multi-head [8]	84.72	88.16	86.4	-	-	-	72.1	72.24	74.58
	BiLSTM + SDP [9]*	82.7	86.7	84.6	-	-	-	67.5	75.8	71.4
	CNN + Global features [10]*	79.5	79.6	79.5	-	-	-	64	62.9	63.4

Performance on ADE. ∗indicates that the corresponding paper does not state if NER performance is micro-average ormacro-average, though we use the micro-average columns for these cases.

Results: Ablation Study

Model	NER			Boundaries RE			Strict RE		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
SpERT.PL (SciBERT)	69.87	71.47	70.66	52.06	51.26	51.65	40.49	39.87	40.18
– **POS embeddings**	69.52	70.66	70.09	51.64	50.82	51.23	39.59	38.95	39.26
– **entity logits**	69.41	70.49	69.96	51.34	49.66	50.48	39.51	38.23	38.86

Conclusion

- Proposed a deep neural model called SpERT.PL for entity and relation extraction from scientific documents.
- POS information and predicted entity logits boost the classification performance.

- Future work:
  - Does dependency parse of the sentences improve classification accuracy?
  - How is the performance on other datasets?
  - How do NER & RE impact other downstream tasks?

References

1. M Eberts, and A Ulges. Span-based Joint Entity and Relation Extraction with Transformer Pre-training. In *EACL*, 2020.
2. I Beltagy, K Lo, and A Cohan. SciBERT: A Pretrained Language Model for Scientific Text. In *EMNLP-IJCNLP*, 2019.
3. J Lee, et. al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36(4), 2020.
4. D Wadden, et al. Relation, and Event Extraction with Contextualized Span Representations. In *EMNLP-IJCNLP*, 2019.
5. Y Luan, et al. A general framework for information extraction using dynamic span graphs. In *NAACL*, 2019.
6. Y Luan, et al. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction. In *EMNLP*, 2018.

References

7. G Bekoulis, et al. Adversarial training for multi-context joint entity and relation extraction. In *EMNLP*, 2018.
8. G Bekoulis, et al. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications* 114, 2018.
9. F Li, et al. A neural joint model for entity and relation extraction from biomedical text. *BMC Bioinformatics* 18(1), 2017.
10. F Li, et al. Joint Models for Extracting Adverse Drug Events from Biomedical Text. In *IJCAI*, 2016.
11. J Wang and W Lu. Two Are Better than One: Joint Entity and Relation Extraction with Table-Sequence Encoders. In *EMNLP*, 2020.
12. S Zhao, et al. Modeling Dense CrossModal Interactions for Joint Entity-Relation Extraction. In *IJCAI*, 2020.
13. Z Zhong and D Chen. A Frustratingly Easy Approach for Entity and Relation Extraction. In *NAACL*, 2021.
14. T Tran and R Kavuluru. Neural metric learning for fast end-to-end relation extraction. *arXiv preprint arXiv:1905.07458*, 2019.
15. Gábor, Kata, et al. Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers. In *SemEval*, 2018.

Thank You