Entity, Relation, and Event Extraction with Contextualized Span Representations

David Wadden, Ulme Wennberg, Yi Luan, Hannaneh Hajishirzi, Paul G.

presented by Alexander Donets

Novosibirsk State University, 2020

Contents

Introduction

Authors present a framework (model and data processing tools) for solving three **Information Extraction** tasks:

- **Named Entity Recognition** (NER)
- **Relation Extraction** (RE)
- **Event Extraction** (EE)

and also they solve auxiliary **Coreference Resolution** (CR) task to enhance model inference.

Introduction

Def. 1

Text Span - a continuous sequence of tokens or their embeddings (vector representations of tokens).

Def. 2

Span Graph - a graph of entities with edges defining coreferences, relations and events.

Problem Definition

Given a document D represented as a sequence of words, we derive $S = \{S_1, ..., S_T\}$, the set of all possible within-sentence text spans.

Then, our goal is to solve:

- CR task: predict antecedent c_i for each span s_i
- NER task: predict entity type e_i for each span s_i,
- RE task: predict relation r_{ij} for each pair of within-sentence spans (s_i, s_j),
- EE task: assign an event trigger label d_i for each predicted entity and predict event arguments a_{ij} for all spans s_j in the same sentence as d_i.

Global Context

One of the toughest problems to address during solving these tasks is the problem of **global context modeling**. To make use of distant links between entities in text is not a trivial problem. Authors try to address this by employing several ideas.

Highlights

What key ideas authors employ in their work?

- Using **Multi-Task Learning**
- Using **pre-trained BERT as text encoder**. BERT creates contextual embeddings and eliminates need for hand-crafted features extraction. Authors test two options of using BERT: for fine-tuning alone, and with biLSTM after BERT.
- Using **Span Graph** to incorporate global context information, that is outside of text window scope of encoder.
- Using of **coreferences between entities** to further enhance global context modeling

Architecture Description

There are 4-layer structure of model inside framework:

- **Token encoding**: several sentences are fed to BERT
- **Span enumeration**: BERT embeddings are enumerated and coreferences are iteratively resolved and updated.
- **Span graph propagation**: iterative inference for relations and propagation of this information, as well as propagation of corefs and update of graph.
- **Multi-task classification**: final inference stage. Re-contextualized span representations or their pairs are supplied to two-layer FFNN scoring function.

Architecture Description

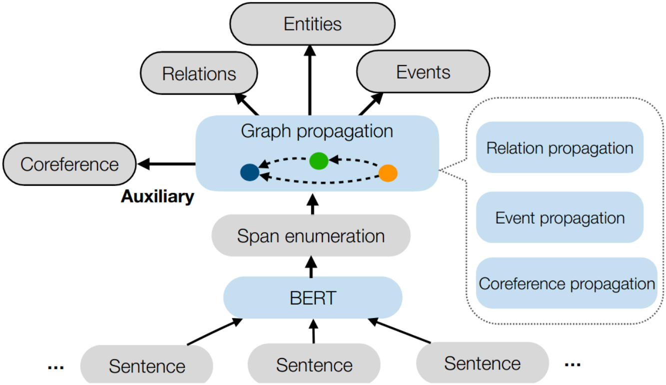

Figure: Overview of DyGIE++ model.

Architecture Description

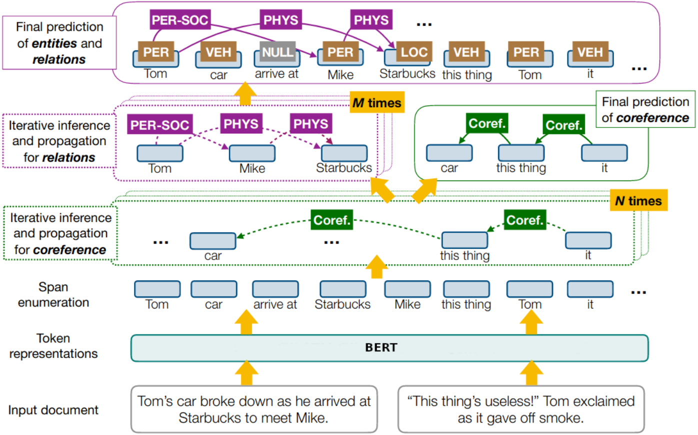

Figure: Overview of DyGIE++ model.

Training

The loss function for training process is defined as weighted sum of log-likelihood for all tasks (because of multi-task learning):

$$\sum_{(D,R^*,E^*,C^*)\in\mathcal{D}} \Big\{ \lambda_{\mathrm{E}} \log P(E^* \mid C, R, D) + \lambda_{\mathrm{R}} \log P(R^* \mid C, D) + \lambda_{\mathrm{C}} \log P(C^* \mid D) \Big\} \quad (5)$$

Datasets

Authors tested their model on several datasets from various domains:

	Domain	Docs	Ent	Rel	Trig	Arg
ACE05	News	511	7	6	-	-
ACE05-E	News	599	7	-	33	22
SciERC	AI	500	6	7	-	-
GENIA	Biomed	1999	5	-	-	-
WLP	Bio lab	622	18	13	-	-

Table 8: Datasets for joint entity and relation extraction and their statistics. *Ent*: Number of entity categories. *Rel*: Number of relation categories. *Trig*: Number of event trigger categories. *Arg*: Number of event argument categories.

Experiments and Results

Dataset	Task	SOTA	Ours	Δ%
ACE05	Entity	88.4	**88.6**	1.7
	Relation	63.2	**63.4**	0.5
ACE05-Event*	Entity	87.1	**90.7**	27.9
	Trig-ID	73.9	**76.5**	9.6
	Trig-C	72.0	**73.6**	5.7
	Arg-ID	**57.2**	55.4	-4.2
	Arg-C	52.4	**52.5**	0.2
SciERC	Entity	65.2	**67.5**	6.6
	Relation	41.6	**48.4**	11.6
GENIA	Entity	76.2	**77.9**	7.1
WLPC	Entity	79.5	**79.7**	1.0
	Relation	64.1	**65.9**	5.0

Figure: DYGIE++ achieves state-of-the-art results. Test set F1 scores of best model, on all tasks and datasets.

Keynotes

- The largest gains from Coreference propagation are on the CS research abstracts of SciERC, which have lots or coreferences, acronyms and abbreviations.
- Relation propagation improves relation extraction performance over pre-trained BERT, but does not improve fine-tuned BERT. Probably, because relations predicted only within same sentence.

Keynotes

- Benefits of cross-sentence context with BERT: model achieved best results across all relation and extraction tasks with 3-sentence window.
- On Event Extraction, BERT fine-tuning decreases performance by 1.6 F1 on average across tasks. Probably, because of task sensitivity to hyperparameters choice - there were cases when trigger detector begun overfitting before the argument detector had finished training.

Conclusion

The authors pushed forward state-of-the-art solutions to three main tasks of Information Extraction. The paper proves that multi-task learning can be reasonable way to asses global context modeling problem.

Thanks for your attention!