Split-NER: Named Entity Recognition via Two Question-Answering-based Classifications

Jatin Arora
Nuro Inc
jarora@nuro.ai

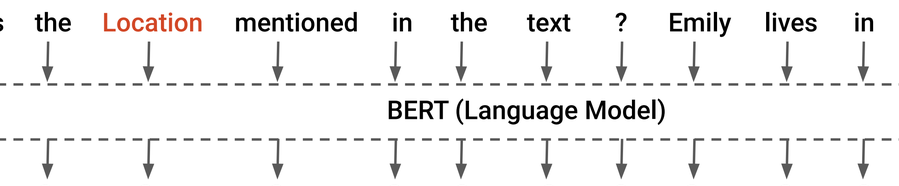
Youngja Park
IBM T. J. Watson Research Center
young_park@us.ibm.com

Source Code: github.com/c3sr/split-ner

Source Code: github.com/c3sr/split-ner

Problem Overview

Named Entity Recognition (NER) is a sub-task of information extraction that aims to locate and classify named entities mentioned in unstructured text.

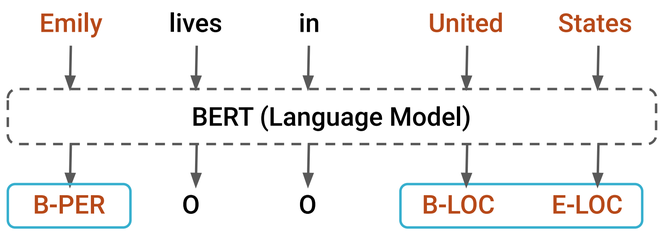


Inherently, there are two tasks involved. However, traditionally, both of these are done together, because they are considered interrelated.

Traditional Approaches

Sequence Tagging

Single(SeqTag)



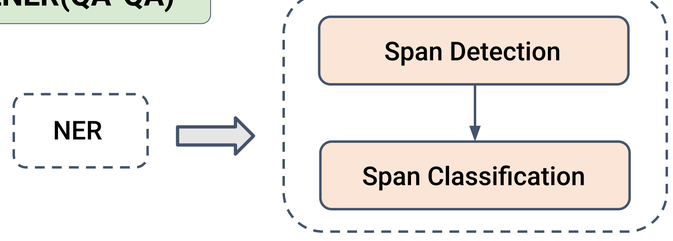
Question-Answering

Single(QA)

Our Approach (Split-NER)

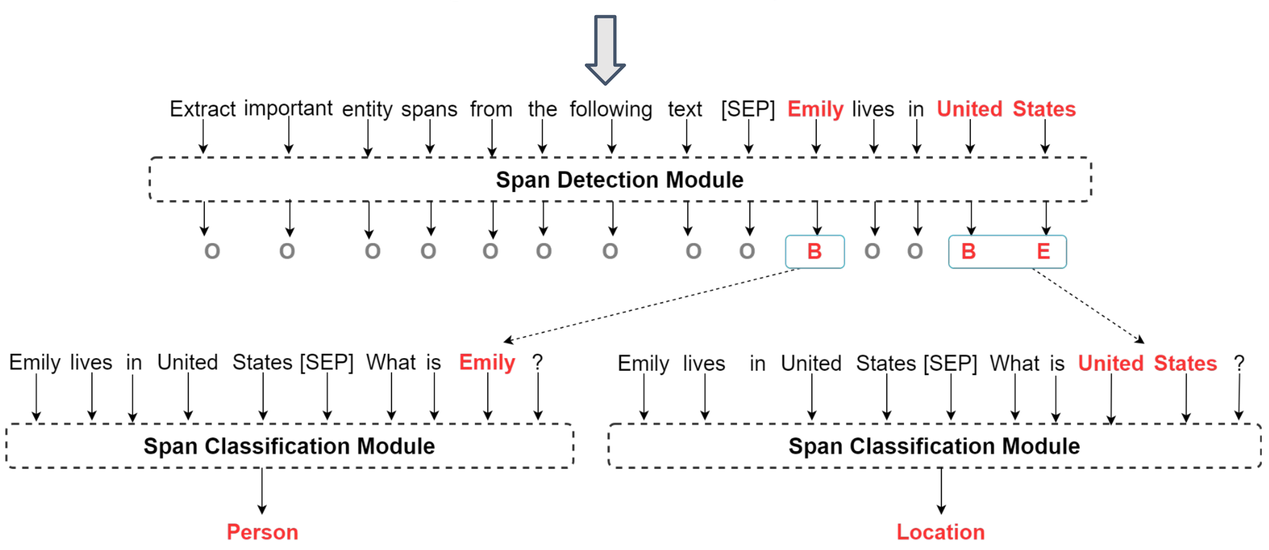
We propose to split and perform NER as a pipeline of two tasks trained independently.

SplitNER(QA-QA)



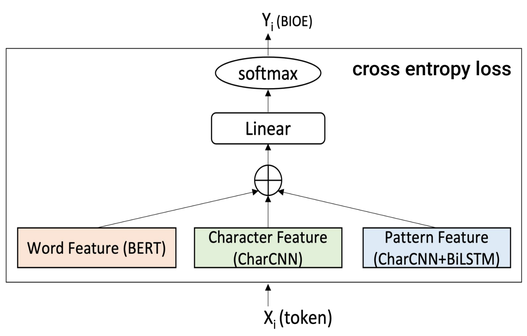
Salient Features

- Performs on-par/better!
- Trains faster!
- Is easily customizable!

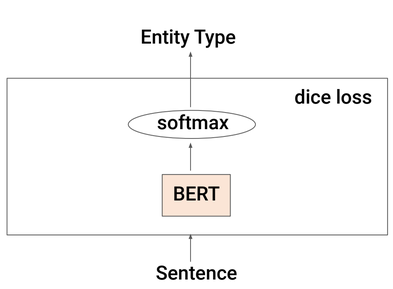


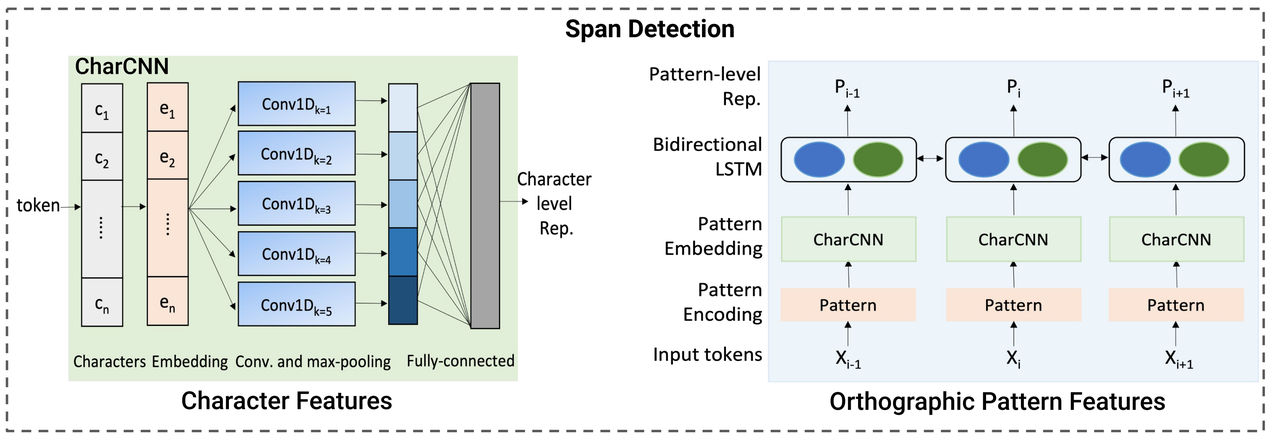
Split-NER Components

Span Detection



Span Classification

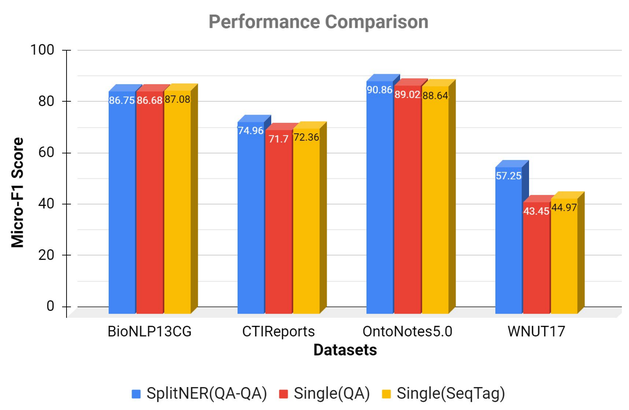


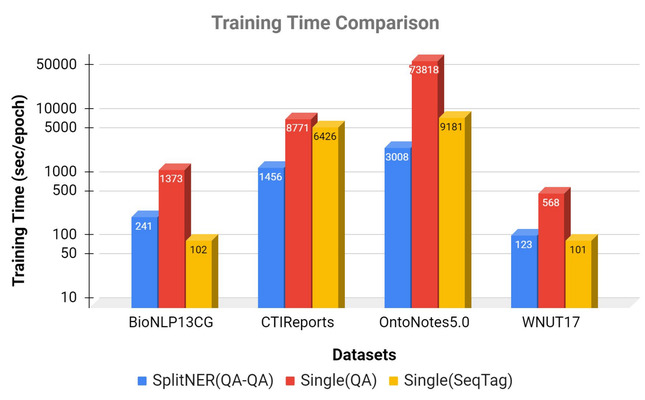


- Pattern Encoding Example: $MgSO_4$ → uluud ← $CaSO_4$
- Bidirectional LSTM helps capture multi-gram patterns.
- During training, Span Classification model takes the ground truth spans.
- During inference, output of Span Detection is fed to Span Classification.

Performance & Training Time

Dataset	Domain	No. of Entities	Dataset Size (~# Sentences)
BioNLP13CG	Science	16	6k
CTIReports	Cyber-Security	8	55k
OntoNotes5.0	News, Conversations	18	77k
WNUT17	Emerging Entities	6	6k





Performance

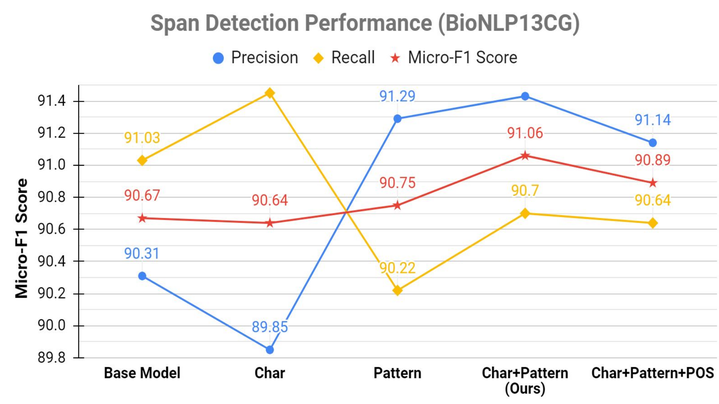
- Split-NER always performs on-par / better than single model approaches.
- On WNUT17, we get a massive 27% improvement compared to baseline.

Training Time

- Split-NER trains on-par / faster than SeqTag and much faster than QA models.
- On OntoNotes5.0, Split-NER trains 25x faster than QA model, 3x faster than SeqTag model.

Span Detection Ablation

Importance of Char & Pattern Features

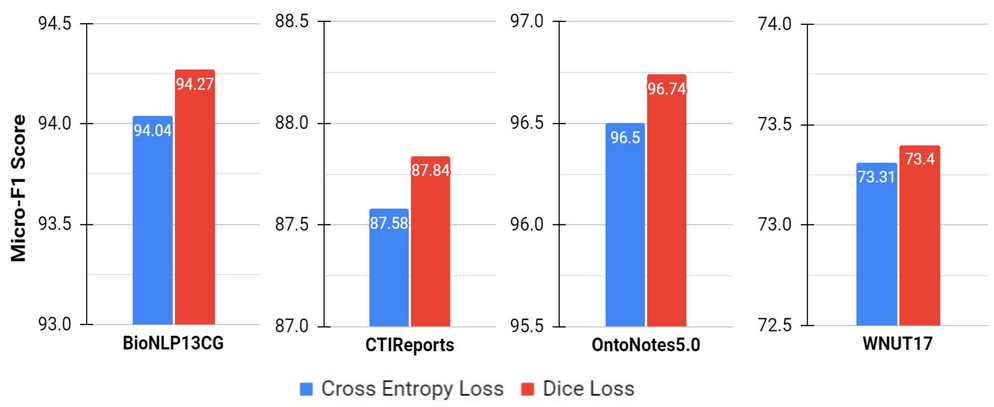


- Char features improve Recall but hurt Precision.
- Pattern features improve Precision but hurt Recall.
- Char+Pattern together give the best Micro-F1.
- POS tags do not help any further.

Sequence Tagging vs Question-Answering

	BioNLP13CG	CTIReports	OntoNotes5.0	WNUT17
SplitNER(QA-QA)	86.75	74.96	90.86	57.25
SplitNER(SeqTag-QA)	86.08	73.84	90.30	56.10

Span Classification Ablation



Class Imbalance

Dice Loss performs better than Cross Entropy Loss.

Qualitative Analysis

Category	Model	Example Sentence
General Detection [Organization]	Single(QA)	*CVS selling their own version of …*
	SplitNER(QA-QA)	*CVS selling their own version of …*
Emerging Entities [Creative Work]	Single(QA)	*Rogue One create a plot hole in Return of the Jedi*
	SplitNER(QA-QA)	*Rogue One create a plot hole in Return of the Jedi*
Scientific Terms [Gene]	Single(QA)	*Treating EU - 6 with anti-survivin antisense …*
	SplitNER(QA-QA)	*Treating EU - 6 with anti-survivin antisense …*
Boundary Fix [Location]	Single(QA)	*Hotel Housekeepers Needed in Spring , TX …*
	SplitNER(QA-QA)	*Hotel Housekeepers Needed in Spring , TX …*
OOV Terms [Product]	Single(QA)	*Store SQL database credentials in a webserver*
	SplitNER(QA-QA)	*Store SQL database credentials in a webserver*
Entity Type Fix [Location -> Product]	Single(QA)	*Why do so many kids in Digimon wear gloves ?*
	SplitNER(QA-QA)	*Why do so many kids in Digimon wear gloves ?*