61 ACL 2023

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

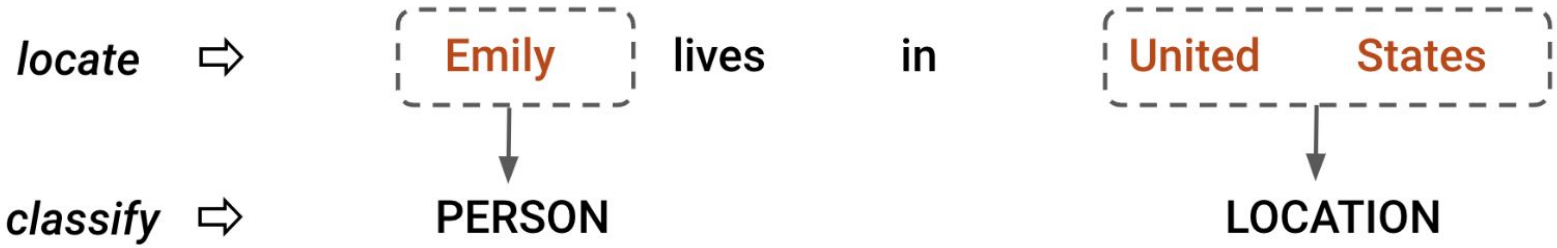
C3SR center for cognitive computing systems research

IBM Research

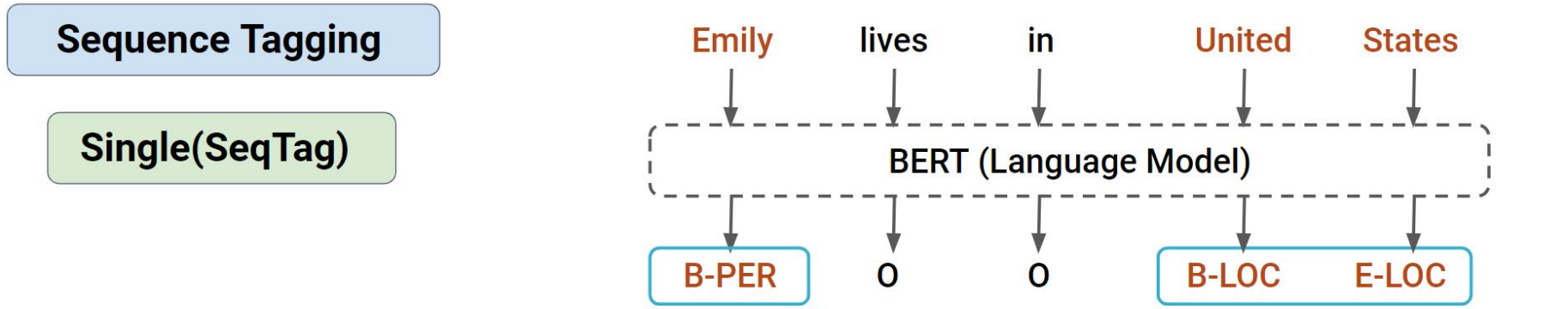
Source Code: github.com/c3sr/split-ner

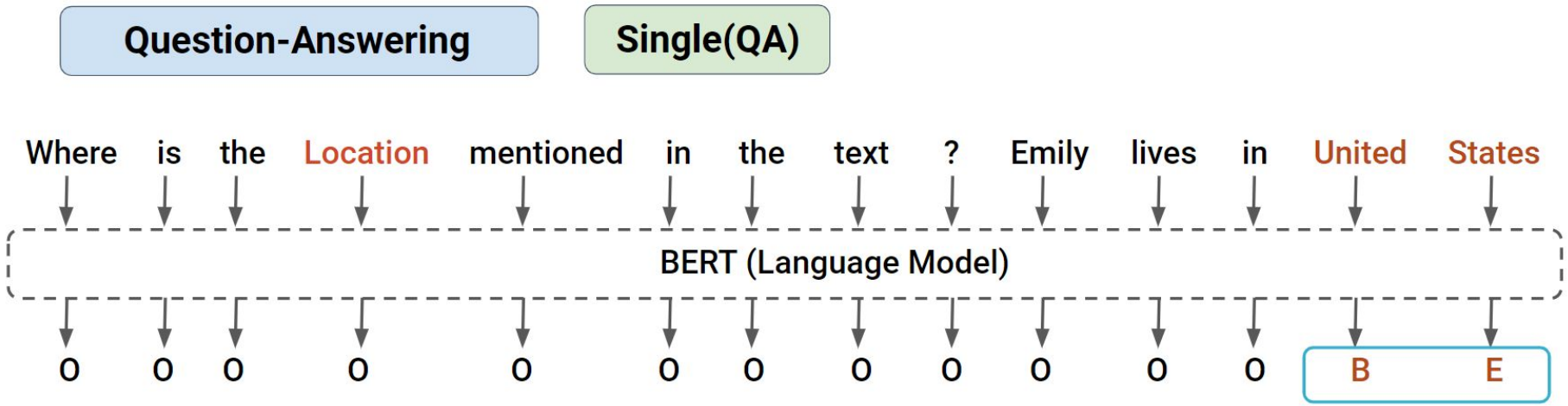
Split-NER Overview

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



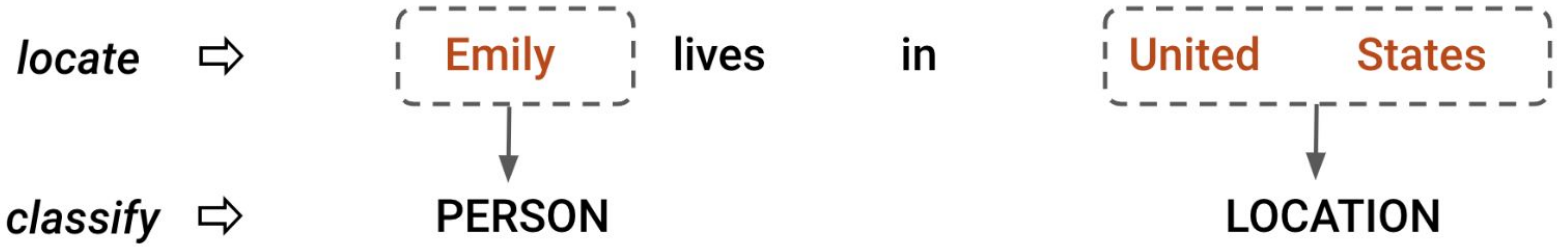
Traditional Approaches

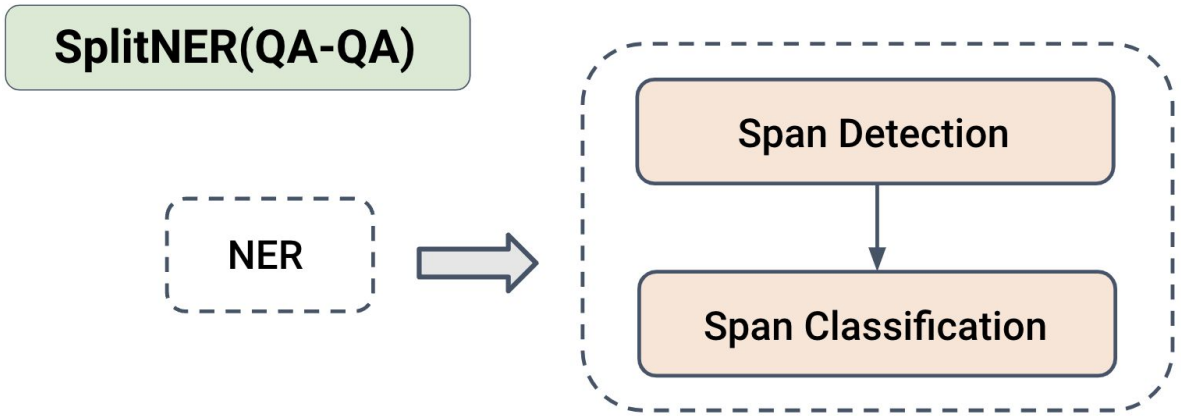




Split-NER Overview

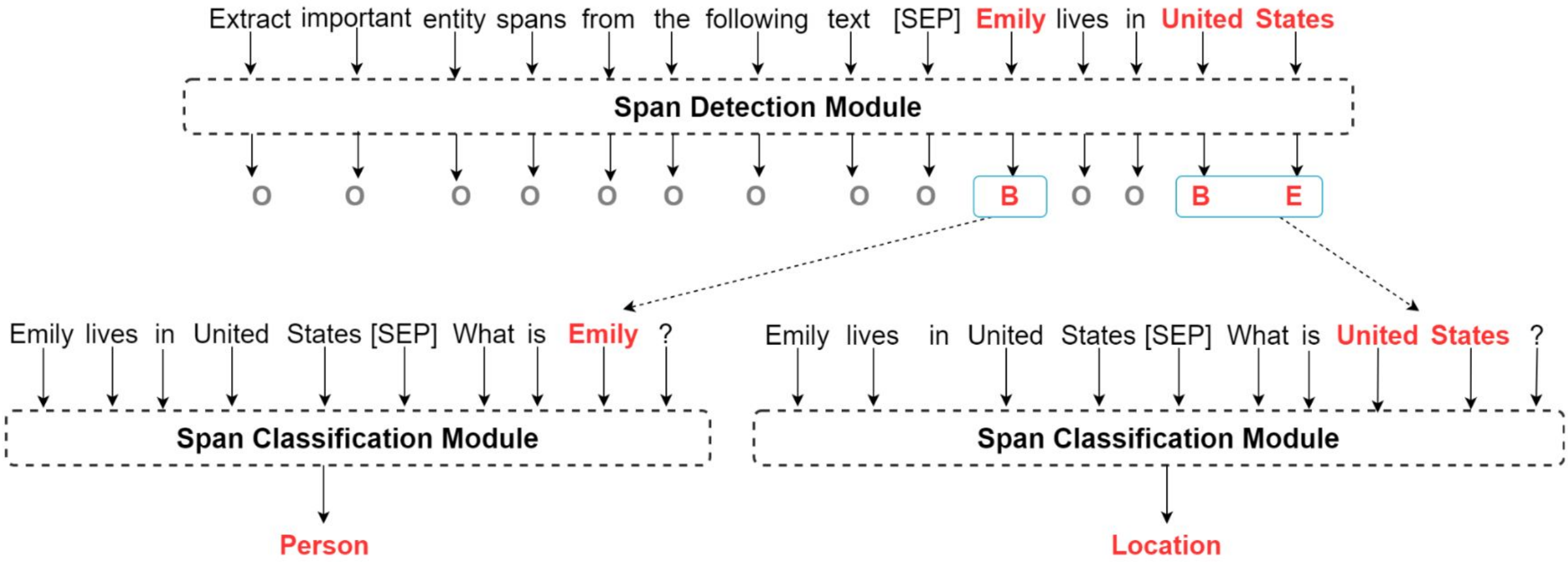
Named Entity Recognition (NER) is a sub-task of information extraction that aims to locate and classify named entities mentioned in unstructured text.





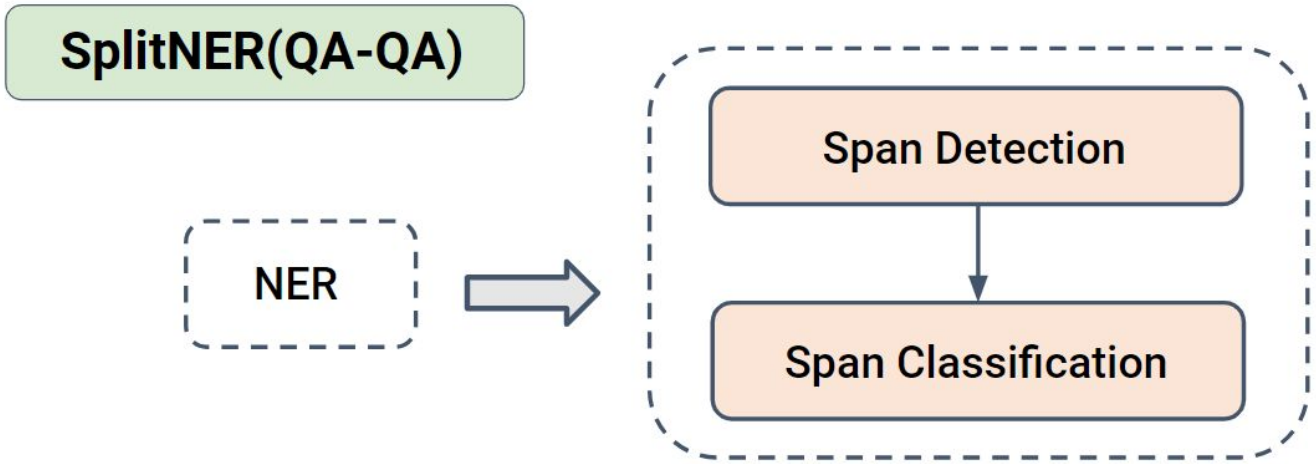
Split-NER Example

Both Span Detection and Span Classification tasks are modeled as QA (Question-Answering) tasks.



Split-NER Overview

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



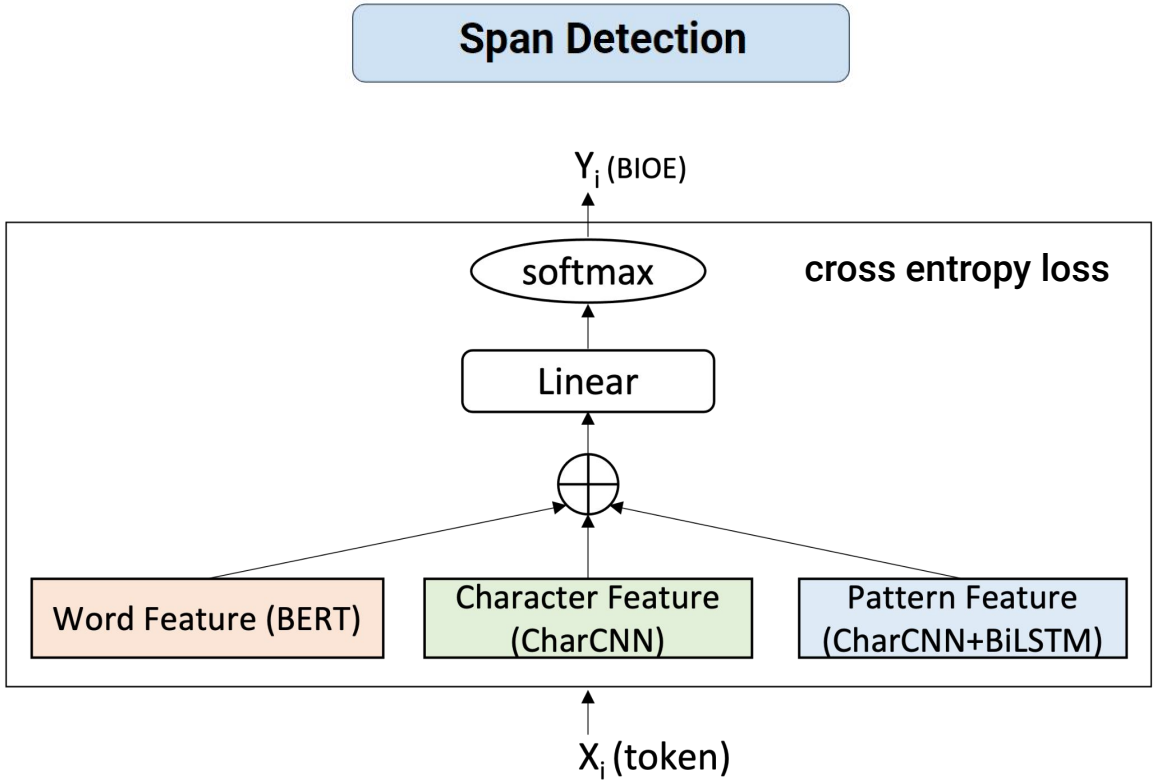
Salient Features

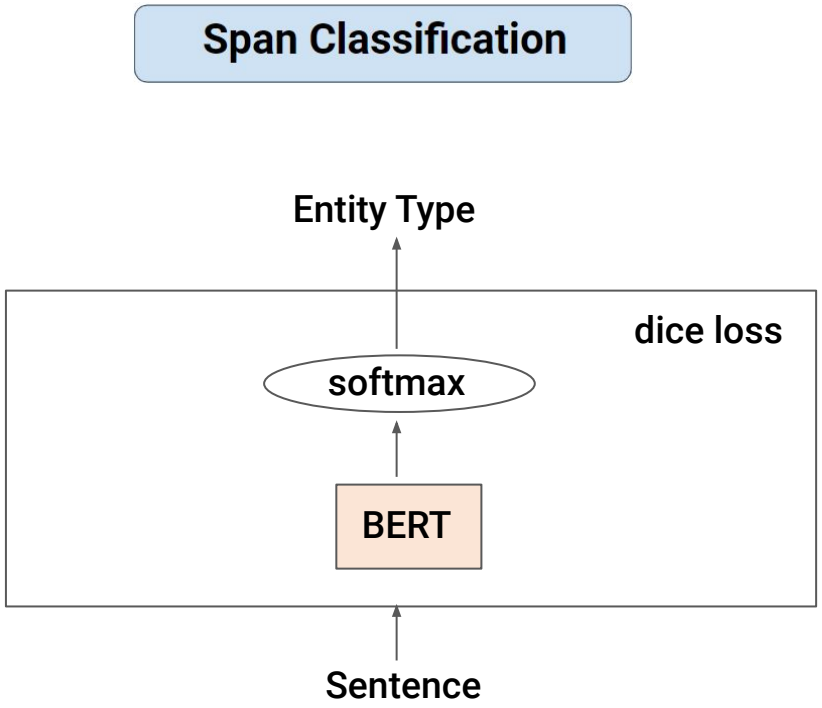
- Performs on-par/better!
- Trains faster!
- Is easily customizable!

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

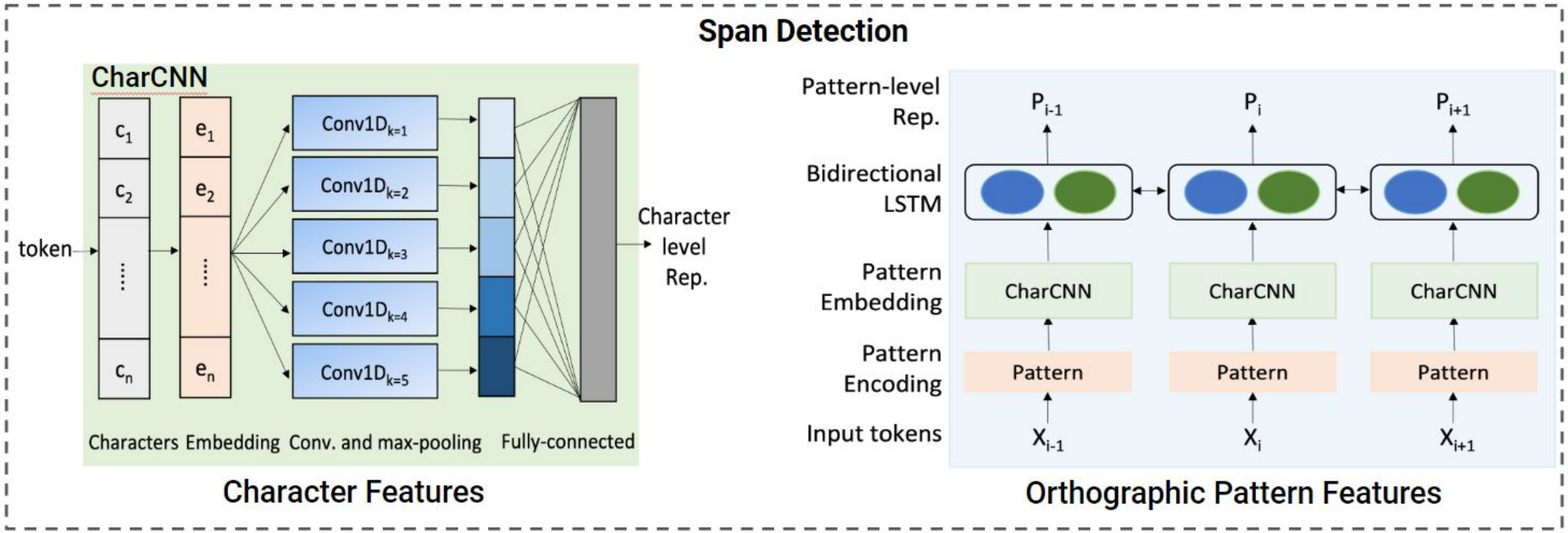
Split-NER Components

- Span Detection Model: Captures BERT semantics + Character embeddings + Orthographic Pattern embeddings.
- Span Classification Model: Does sentence classification. Uses Dice Loss to handle class imbalance.



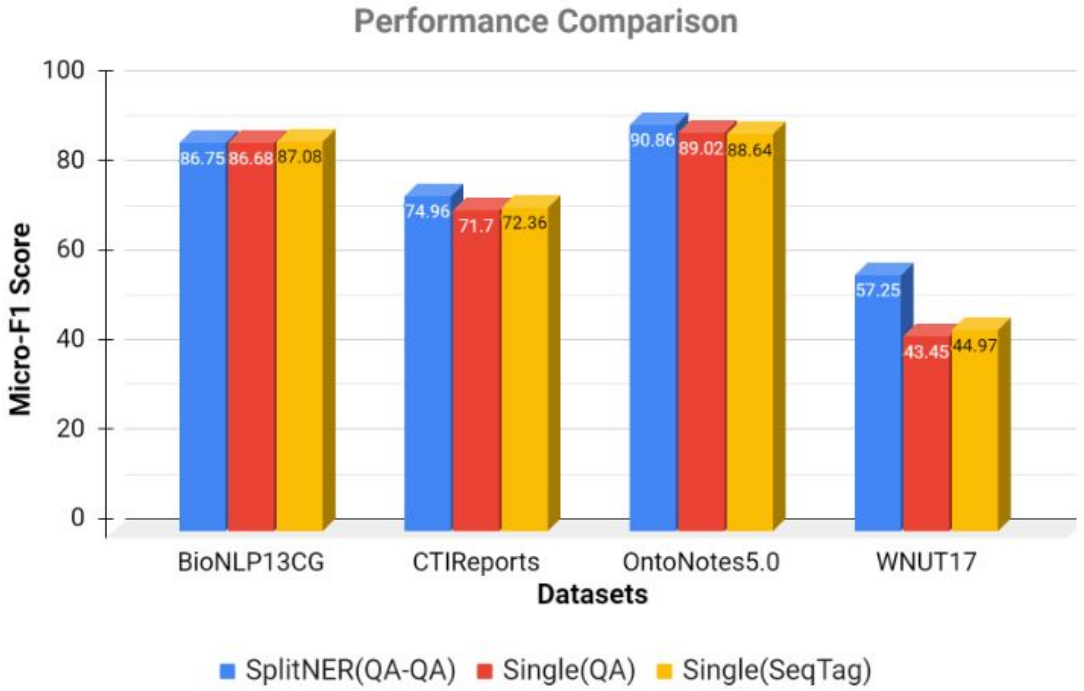


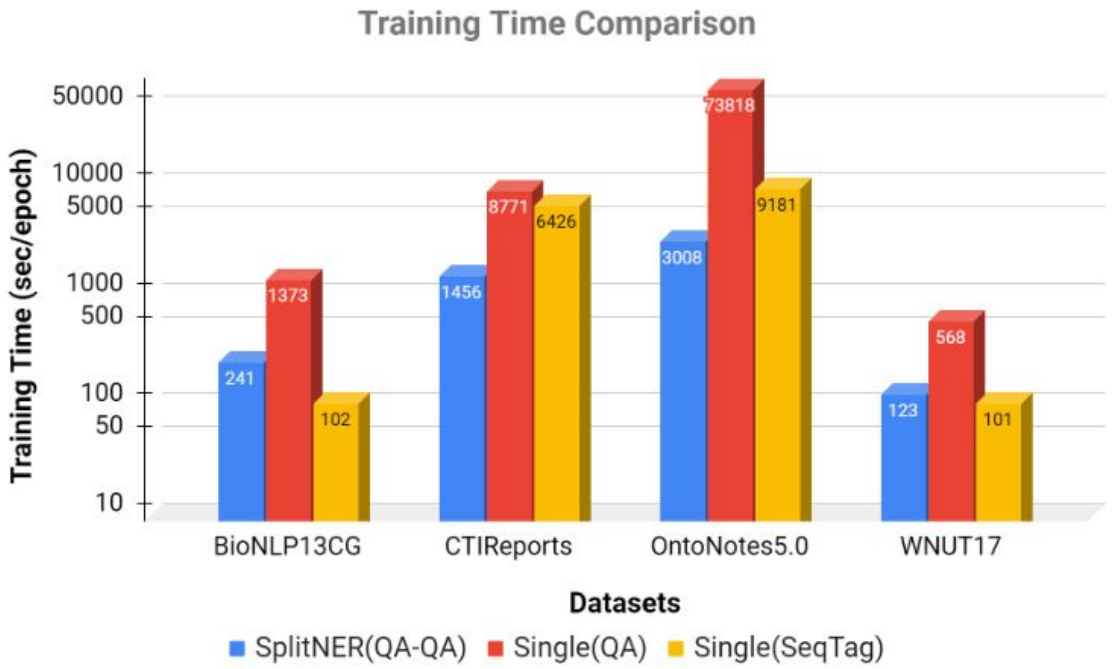
Span Detection Model (Char + Pattern Features)



- 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.

Results: Performance & Training Time





Performance

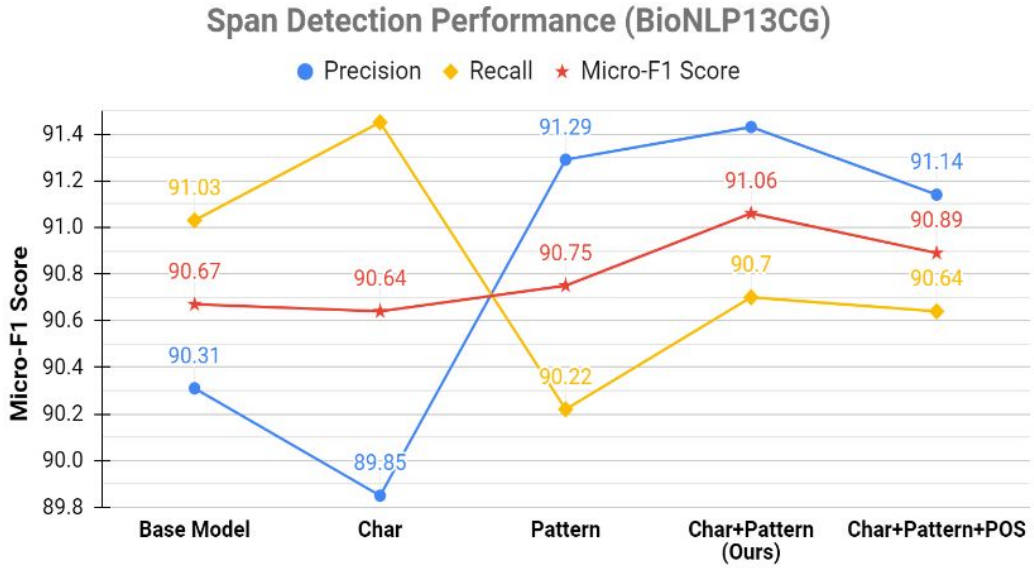
- Split-NER always performs on-par / better than single model approaches.
- On WNUT17, we get a massive 27% improvement compared to baseline.

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

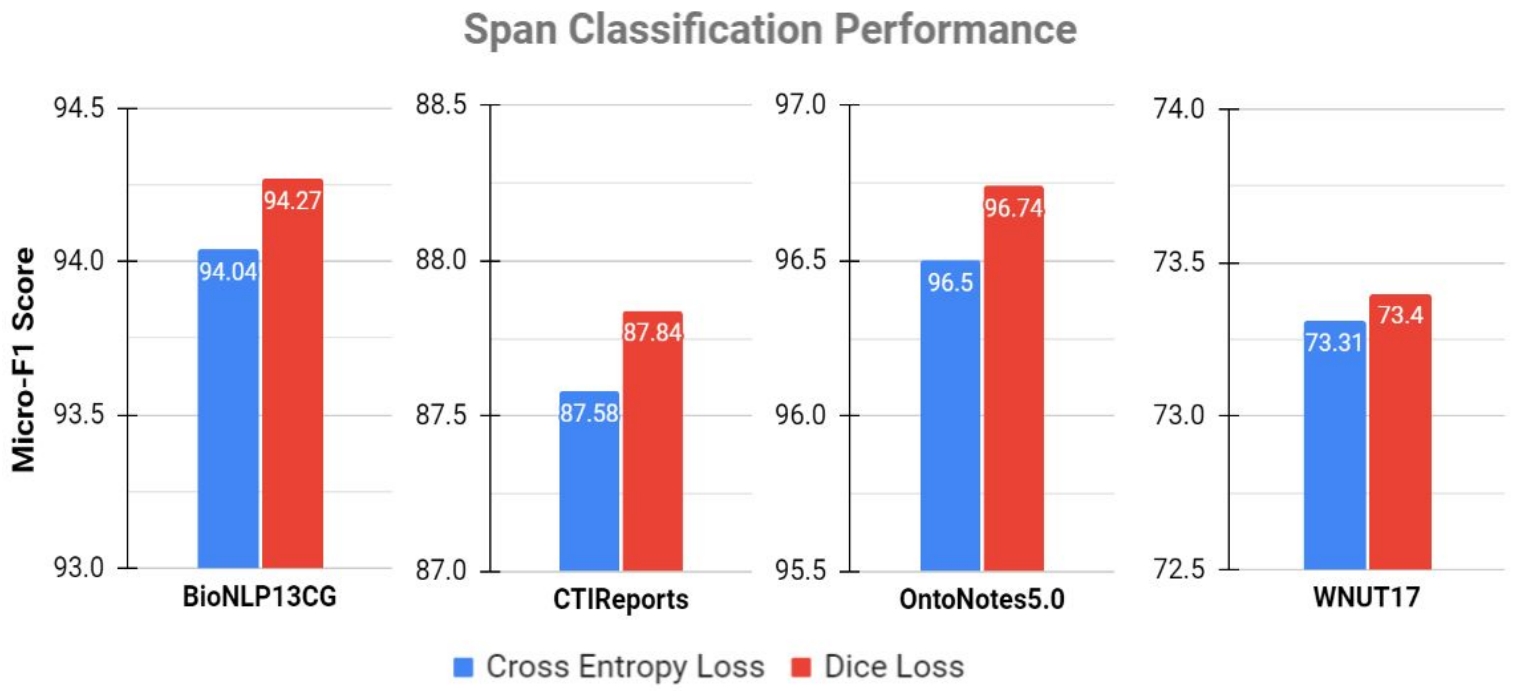


- Char features improve Recall but hurt Precision.
- Pattern features improve Precision but hurt Recall.
- Char+Pattern together give the best Micro-F1.
- POS tags do not help any further.

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 (Loss Function)



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 ?*

Learnings & Key Takeaways

- NER task can be split into two independently trainable tasks (Split-NER):
  - Span Detection
  - Span Classification

- Experiments across 4 cross-domain datasets show that Split-NER:
  - Performs on-par / better!
  - Trains faster!
  - Is easily customizable!
    - Char + Pattern features (Span Detection)
    - Ex: Dice Loss (Span Classification)

Source Code and pre-trained model checkpoints: github.com/c3sr/split-ner

Thank You!