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

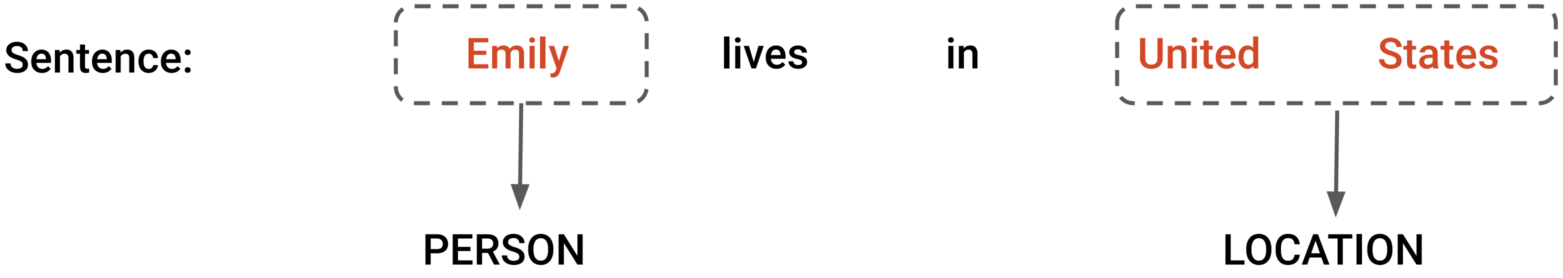
Source Code: github.com/c3sr/split-ner

Problem Overview

This paper belongs to the domain of Named Entity Recognition (NER).

As per Wikipedia,

Named Entity Recognition (NER) is a sub-task of information extraction that seeks to locate and classify named entities mentioned in unstructured text into predefined categories like person names, organizations, locations etc.



Sentence: Emily lives in United States

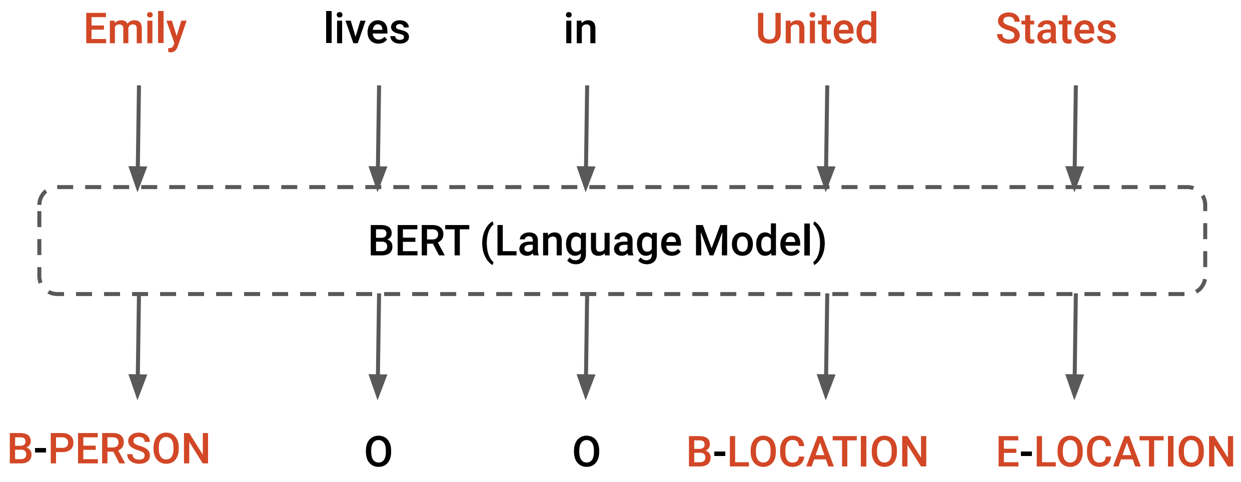
PERSON

LOCATION

In this paper, we propose to split up NER task into subtasks for first locating and then classifying names entities.

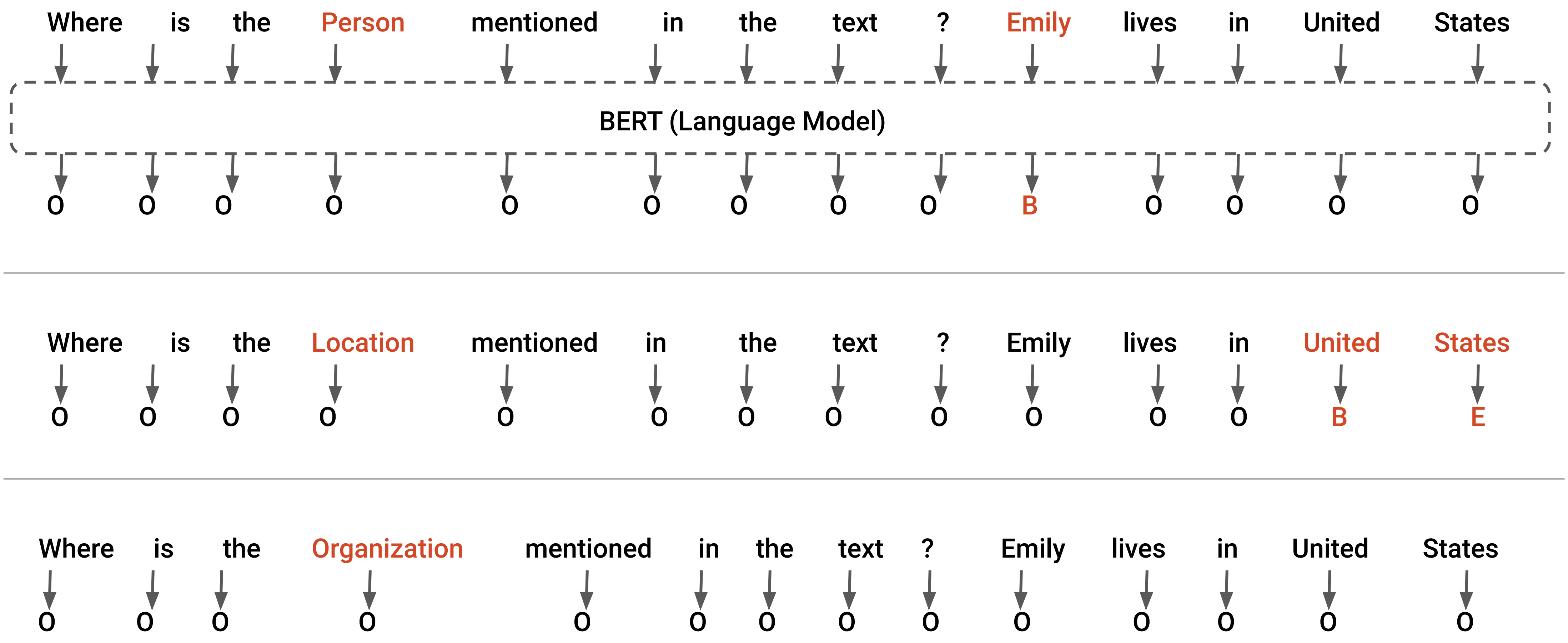
Traditional Approaches (Sequence Tagging - SeqTag)

- Traditionally, locating and classifying named entities are seen as correlated tasks and done together.
- Sequence Tagging: classify each token to an entity type, trained using classification loss.
- Multi-gram entities are captured using BIOE tagging scheme.
- Each sentence is fed only once. Input Complexity: O(N)
  - N: count of sentences
  - T: count of entity types



Traditional Approaches (Question Answering - QA)

Feeds entity type as a question. Each sentence and entity type pair is fed to the model. Input Complexity: O(NT)



Where is the Person mentioned in the text ? Emily lives in United States

BERT (Language Model)

O O O O O O O O O B O O O O

Where is the Location mentioned in the text ? Emily lives in United States

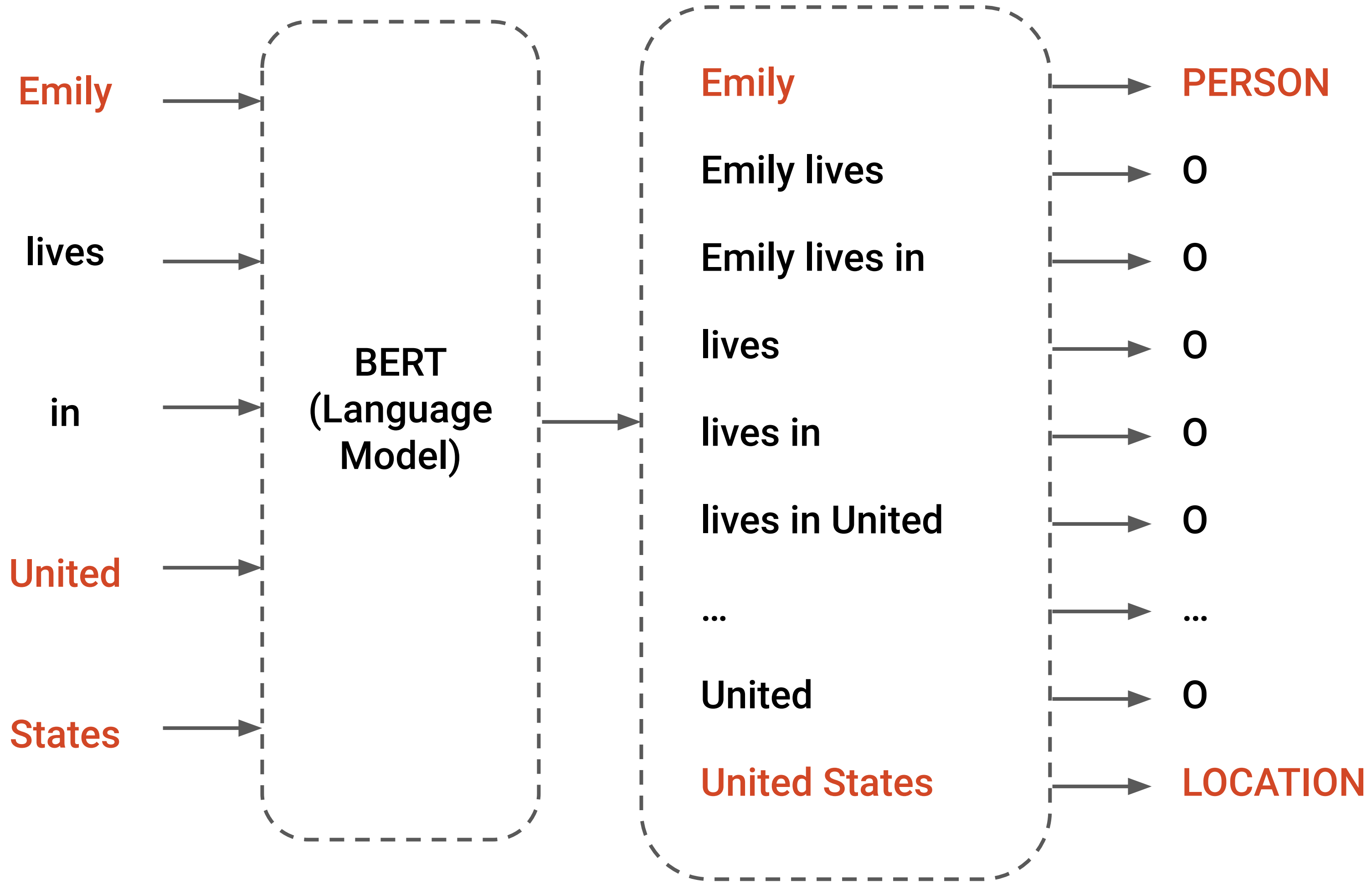
O O O O O O O O O O O O B E

Where is the Organization mentioned in the text ? Emily lives in United States

O O O O O O O O O O O O O O

Traditional Approaches (Span-based)

Classifies every possible span in the sentence. Modeling complexity: $O(N^2)$.

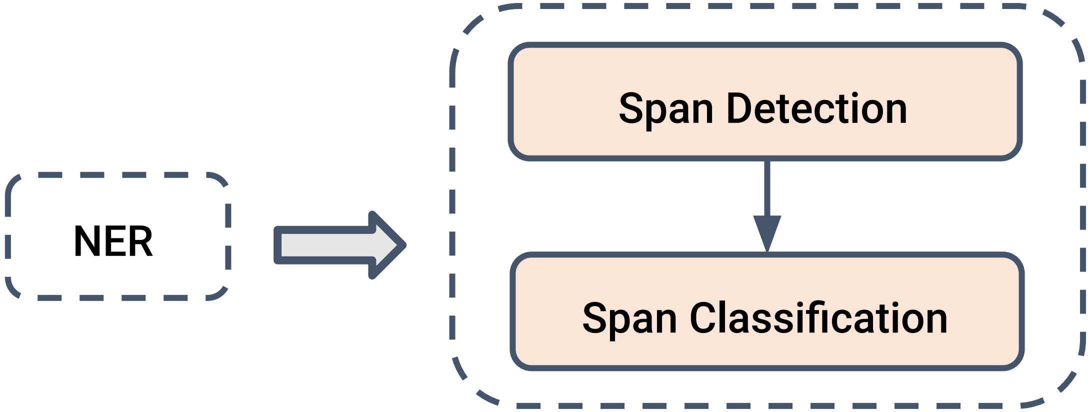


Our Approach (Split-NER)

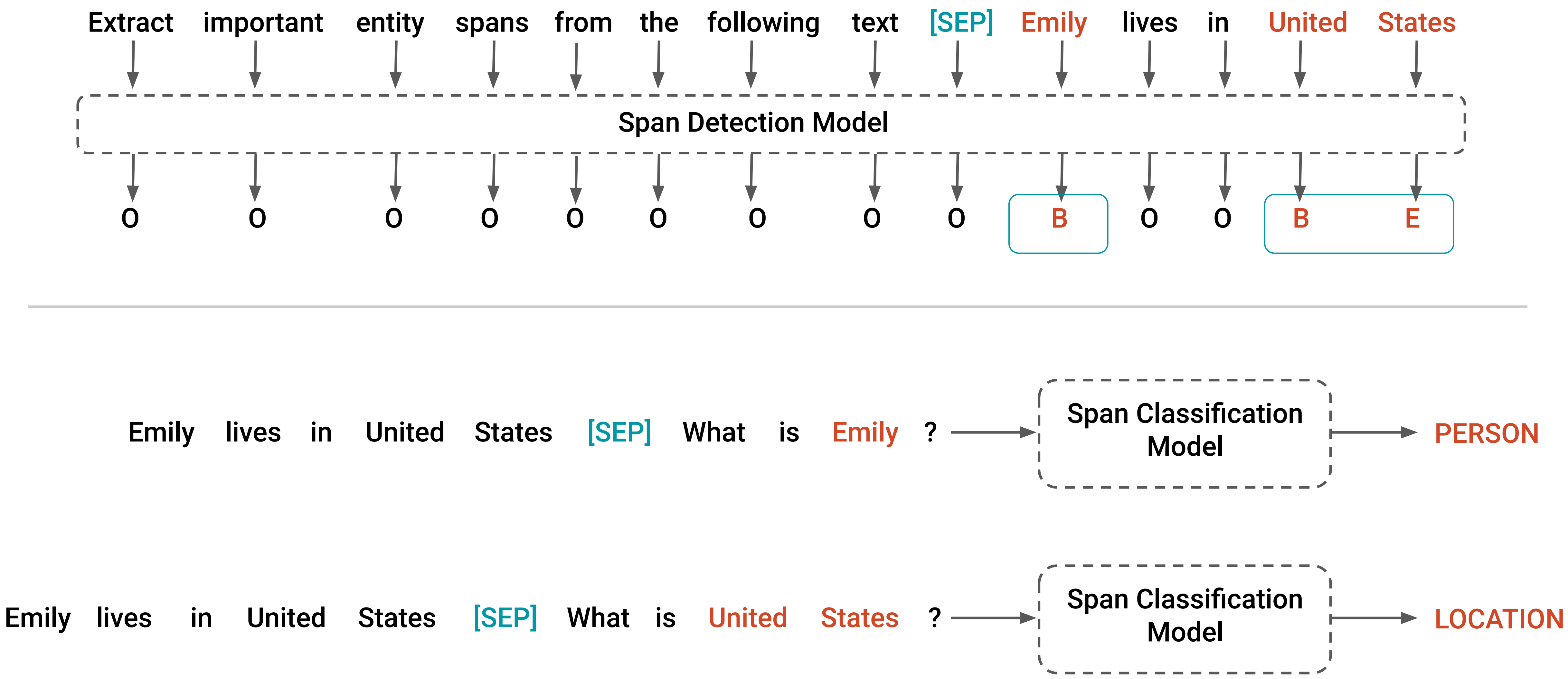
- All of the previous approaches shown treat NER as an atomic task.

- We propose to split NER task into a pipeline of two independently trained sub-tasks:
  - Span Detection
  - Span Classification

- Split-NER approach:
  - Trains faster!
  - Gives matching / better performance!
  - Is easily customizable!



Our Approach (Split-NER)



Extract important entity spans from the following text [SEP] Emily lives in United States

Span Detection Model

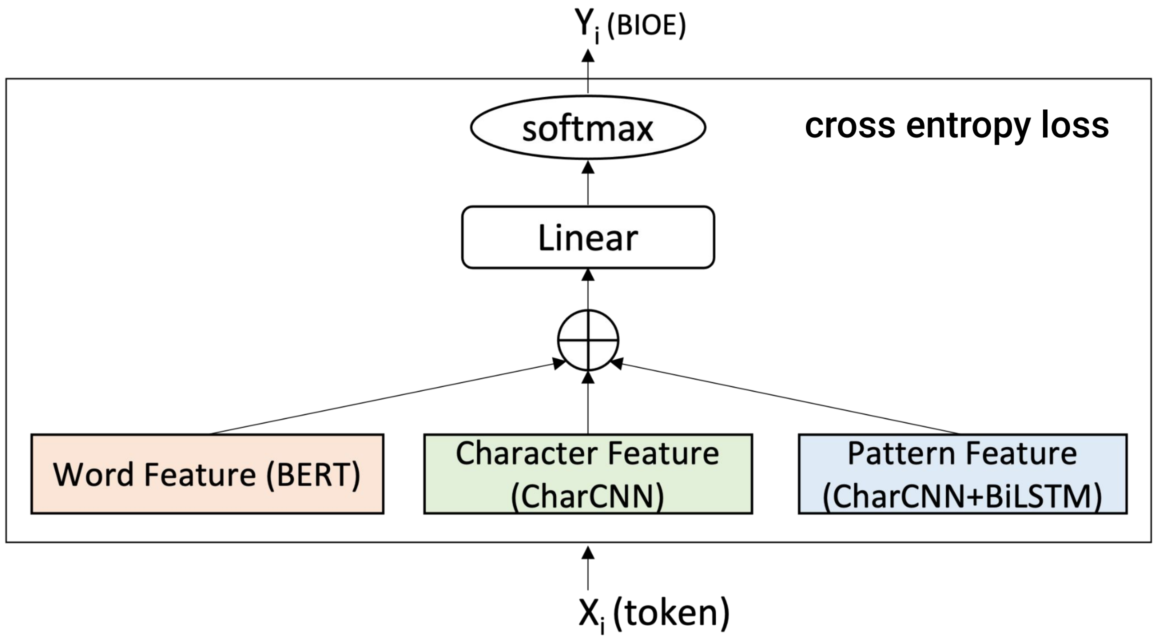
O O O O O O O O O B O O B E

Emily lives in United States [SEP] What is Emily ? → Span Classification Model → PERSON

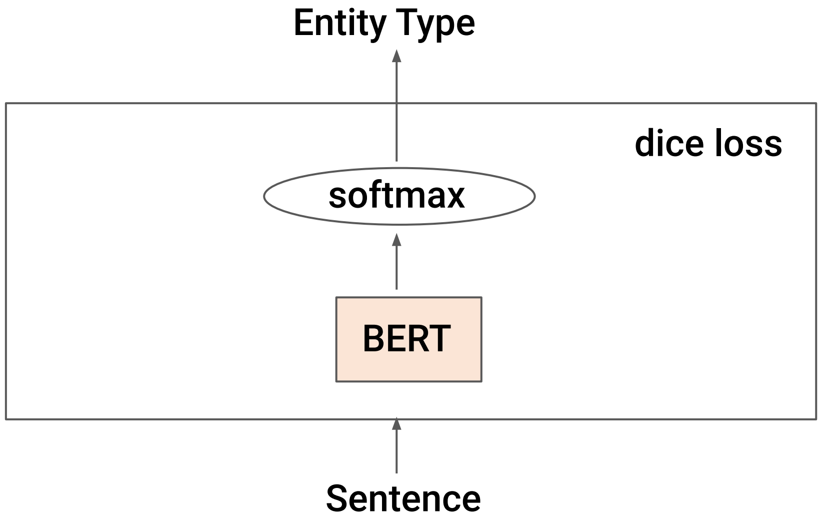
Emily lives in United States [SEP] What is United States ? → Span Classification Model → LOCATION

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.

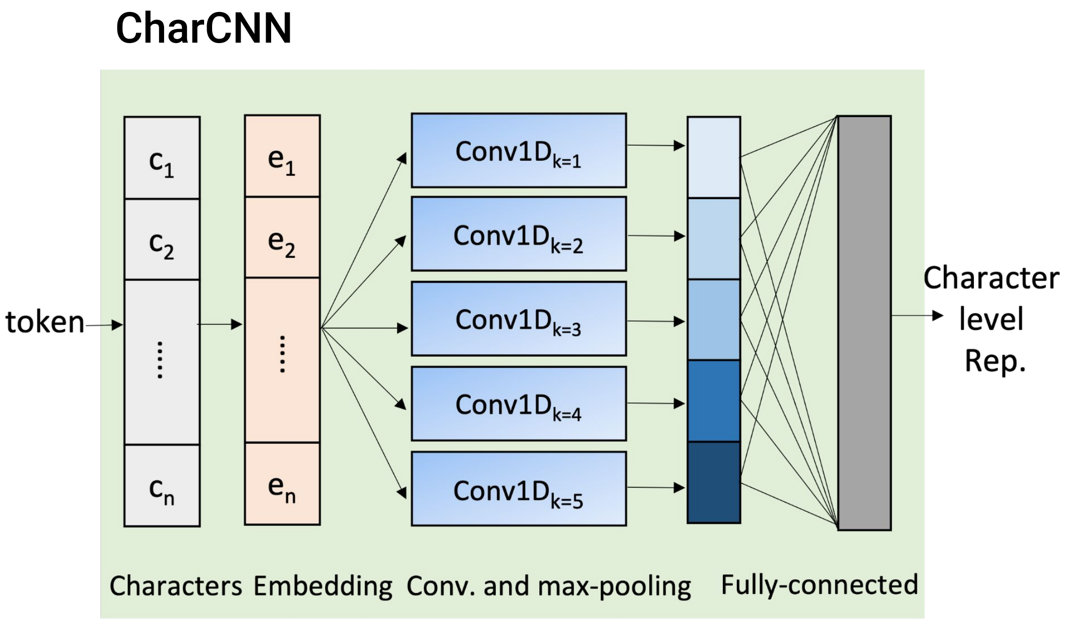


Span Detection Model

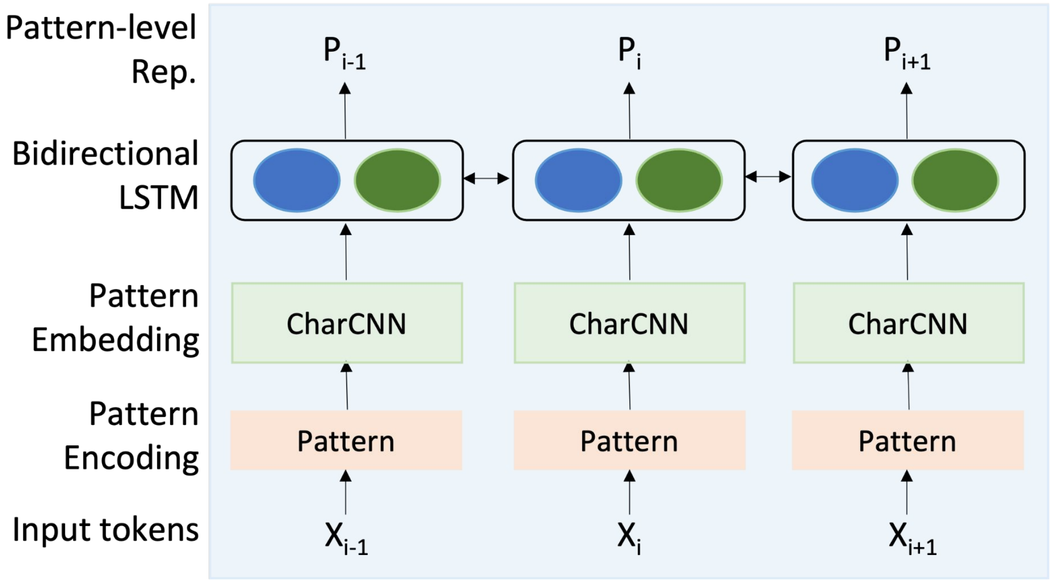


Span Classification Model

Span Detection Model (Char + Pattern Features)



Character Features



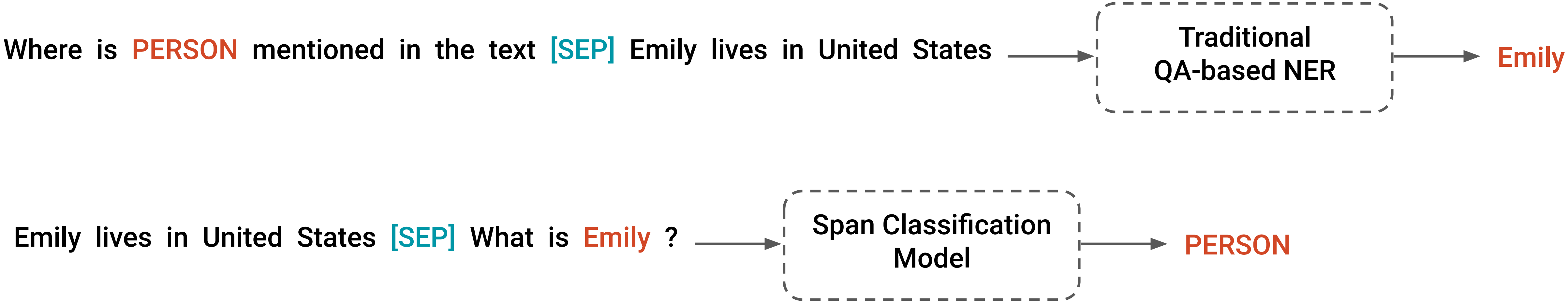
Orthographic Pattern Features

Pattern Encoding Example: $MgSO_4$ → uluud ← $CaSO_4$

Bidirectional LSTM helps capture multi-gram patterns.

Span Classification

- During training, Span Classification takes the ground truth spans as input.
- During inference, Span Classification takes the output of Span Detection as input.
- Span Classification can be seen as the reverse of QA-based NER:
    - QA-based NER: takes entity type as input and outputs spans
    - Span Classification: takes span as input and outputs entity type



Where is PERSON mentioned in the text [SEP] Emily lives in United States → Traditional QA-based NER → Emily

Emily lives in United States [SEP] What is Emily ? → Span Classification Model → PERSON

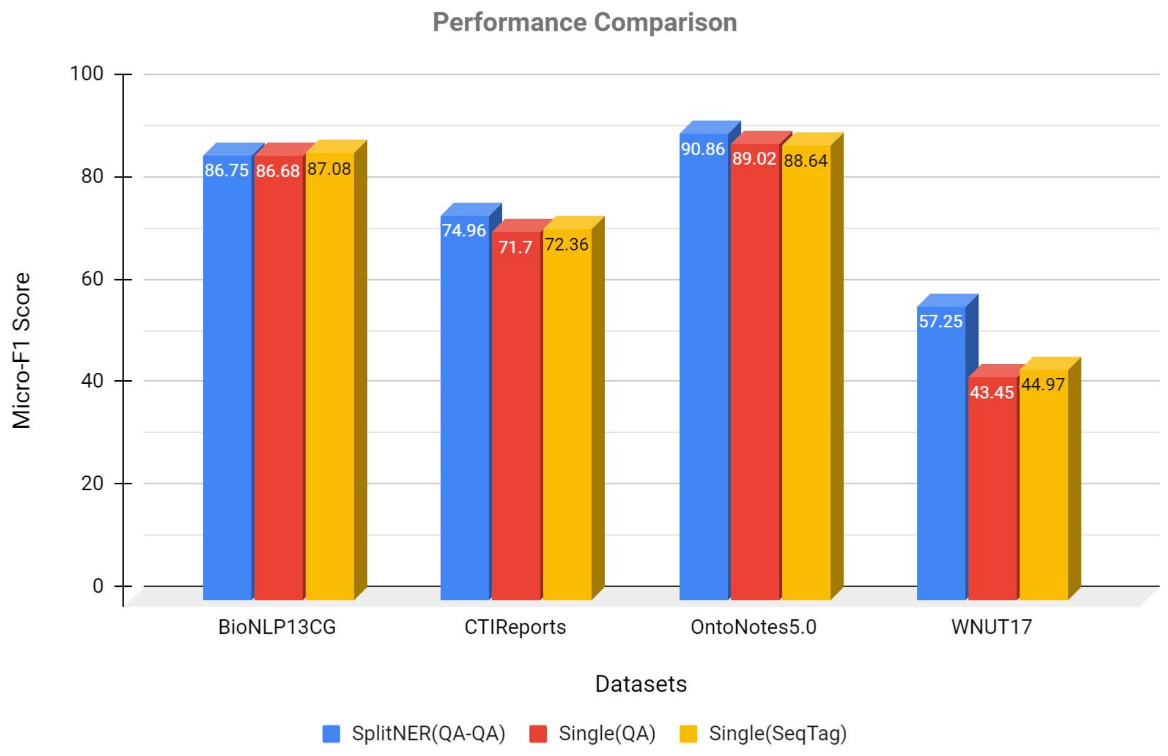
Experimental Setup + Datasets

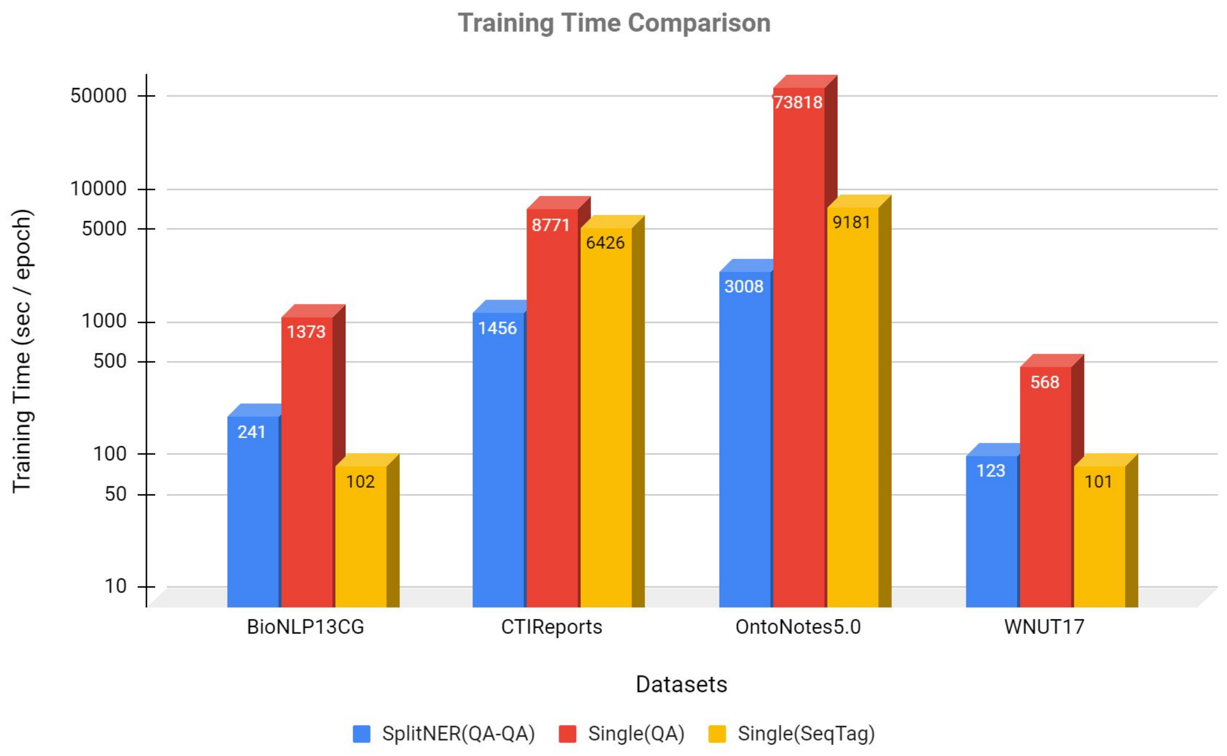
- We use BERT-Base architecture in all experiments.
- We conduct experiments on 4 cross-domain datasets.

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

Model	Definition
SplitNER(QA-QA)	[Our approach] Span Detection (QA) + Span Classification (QA)
Single(QA)	[Baseline] Traditional single NER model (QA)
Single(SeqTag)	[Baseline] Traditional single NER model (Sequence Tagging)

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 Sequence Tagging and much faster than QA-based approach.

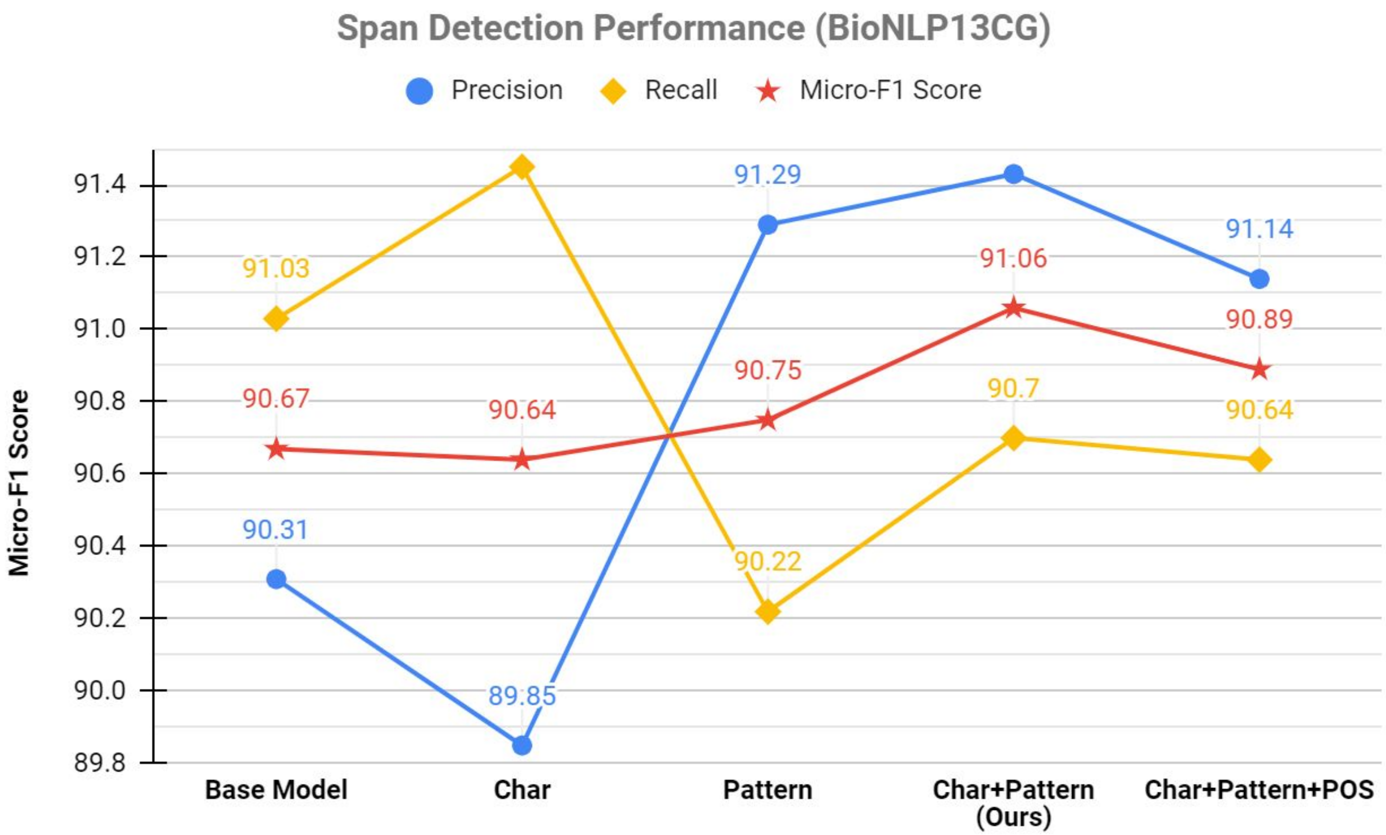
Ablation: Split-NER Variants

- Split-NER is a framework. We also compare by using different techniques for Span Detection and Span Classification.
- For Span Detection, we compare QA-based model with Sequence Tagging (SeqTag) model.
- General Performance Trend: SplitNER(QA-QA) > SplitNER(QA$_{NoCharPattern}$-QA) > SplitNER(SeqTag-QA)

	BioNLP13CG	CTIReports	OntoNotes5.0	WNUT17
SplitNER(QA-QA)	**86.75**	**74.96**	**90.86**	**57.25**
SplitNER(QA$_{NoCharPattern}$-QA)	86.70	74.05	90.58	56.24
SplitNER(SeqTag-QA)	86.08	73.84	90.30	56.10

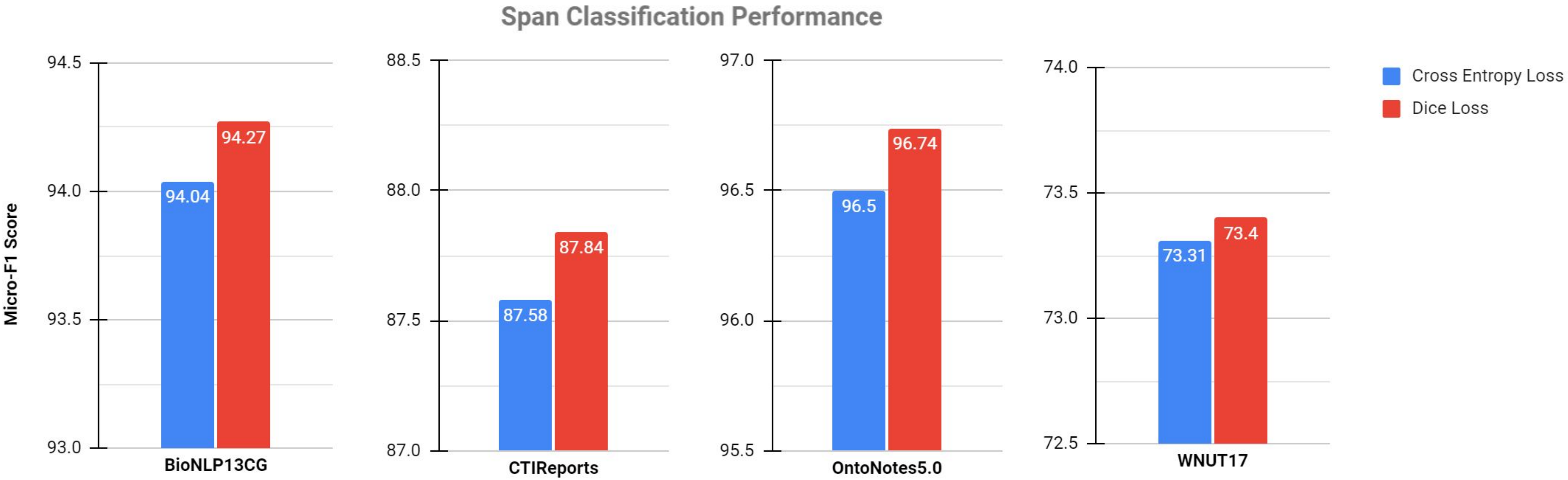
Ablation: Span Detection (Char/Pattern Variations)

- Char features help improve Recall but hurt Precision.
- Conversely, Pattern features help improve Precision but hurt Recall.
- Together, Char+Pattern features give the best Micro-F1.
- Adding additional Part-of-Speech (POS) tags is unnecessary.



Ablation: Span Classification (Loss Function Variation)

- Across all datasets, Dice Loss performs better than Cross Entropy Loss and is able to handle class-imbalance across entity types better.



Qualitative Analysis

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

Learnings & Key Takeaways

- NER task can be split into two independent tasks performed sequentially (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!