Multi-Domain Named Entity Recognition with Genre-Aware and Agnostic Inference

Bloomberg

Engineering

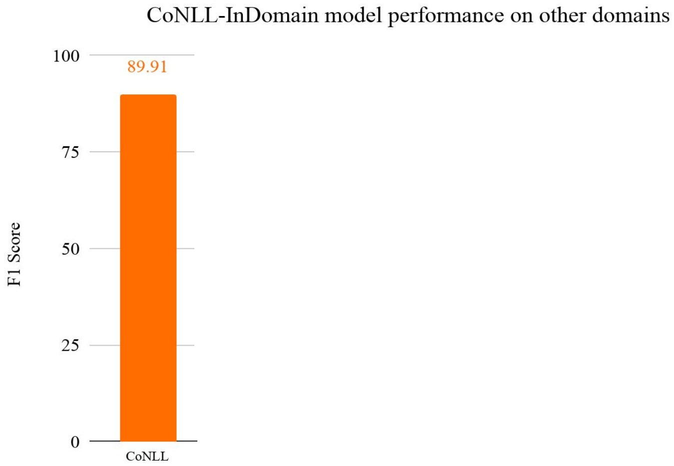
The 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)

Jing Wang, Mayank Kulkarni, Daniel Preoțiuc-Pietro
Research Scientists, AI Group

TechAtBloomberg.com

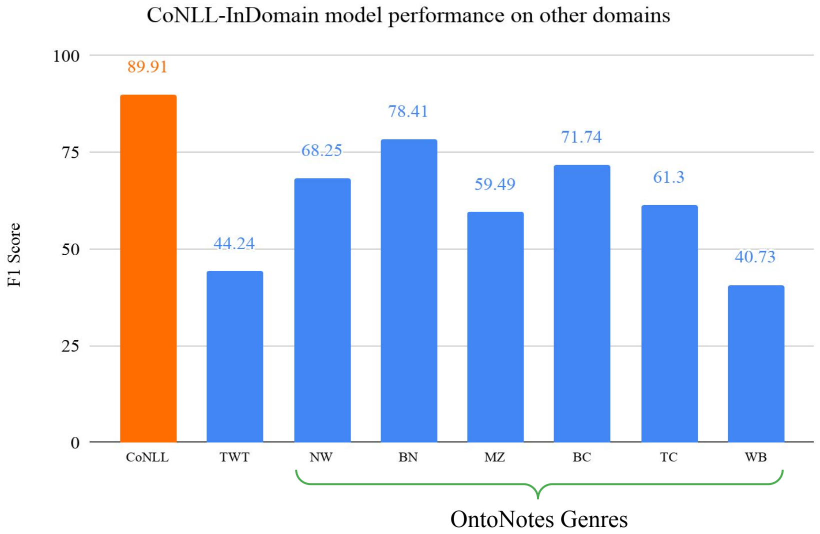
Motivation

- NER models achieve high performance when tested on data from same domain



Motivation

- NER models achieve high performance when tested on data from same domain
- … but generalize poorly on data from other domains

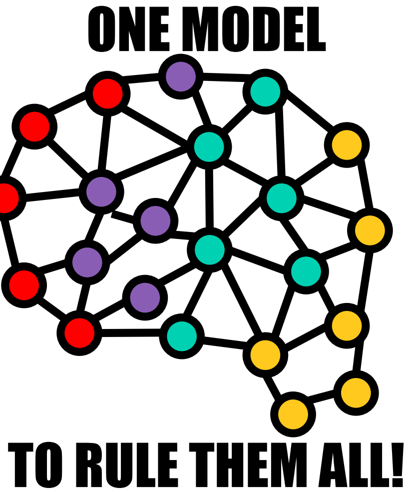


TechAtBloomberg.com

Bloomberg
Engineering

Multi-Domain NER

- Ideally, one model should be able to generalize to multiple domains
  - Traditionally, domain adaptation is posed as single-source, single-target
  - Transitioning to multiple-target domains is non-trivial



Benefits of Multi-Domain NER

- Leverage commonalities across multiple domains
- Leverage specific information for a single domain
- Better generalization to other domains
- Simplifies model maintenance

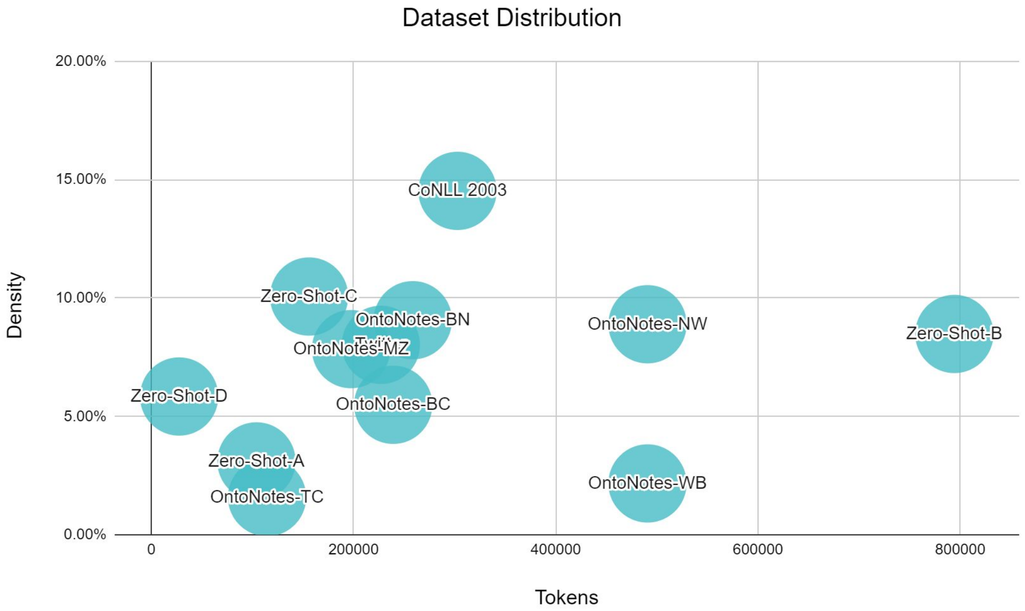
… when compared to training independent models

Experimental Setups

1. Multi-domain with known domain information
    - User provides data from a domain used in training, and knows the domain label
2. Multi-domain with unknown domain information
    - User provides data from a domain used in training, but does not know the domain label
3. Zero-Shot domain
    - User provides data from a completely different domain (i.e., not in training)

Datasets

- **CoNLL 2003** - News Articles
- **Twitter** - 22K English Tweets
    - Temporally-informed Analysis of Named Entity Recognition (ACL 2020)
- **OntoNotes** (6 genres)
    - **NW** - newswire
    - **BN** - broadcast news
    - **MZ** - magazine
    - **BC** - broadcast conversation
    - **TC** - telephone conversation
    - **WB** - web data
- **Zero-shot documents** (4 genres)*
    - **Zero-shot-A**
    - **Zero-shot-B**
    - **Zero-shot-C**
    - **Zero-shot-D**



* Data only used in testing

Base Model Architecture

BiLSTM-CRF [Lample et al., 2016]

- Word Embeddings: GloVe + FastText
- Character Embeddings: Randomly Initialized

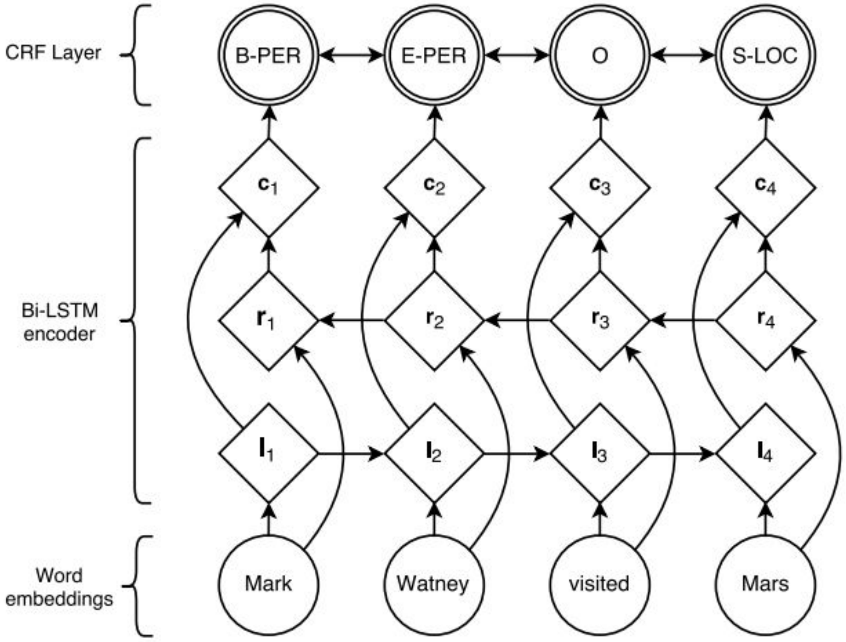


Figure from [“Neural Architectures for Named Entity Recognition”]

Bloomberg Engineering

Baseline Models - InDomain

- Train one NER model for each domain
- Methods:
  - InDomain
  - InDomain + DomainClassifier

    - Train a domain classifier that will determine what in-domain model to run
- Drawbacks:
  - Poor generalization
  - Overhead of add new genres

Baseline Models - PoolDomain

- Global model by pooling data from all genres
- Methods:
    - PoolDomain
    - PoolDomain with INIT Strategy: Initialize weights from learned domain and fine-tune
    - PoolDomain with Gradient Reversal Layer
    - PoolDomain by learning domain embedding with word embeddings
- Drawbacks:
    - Discard genre-specific information

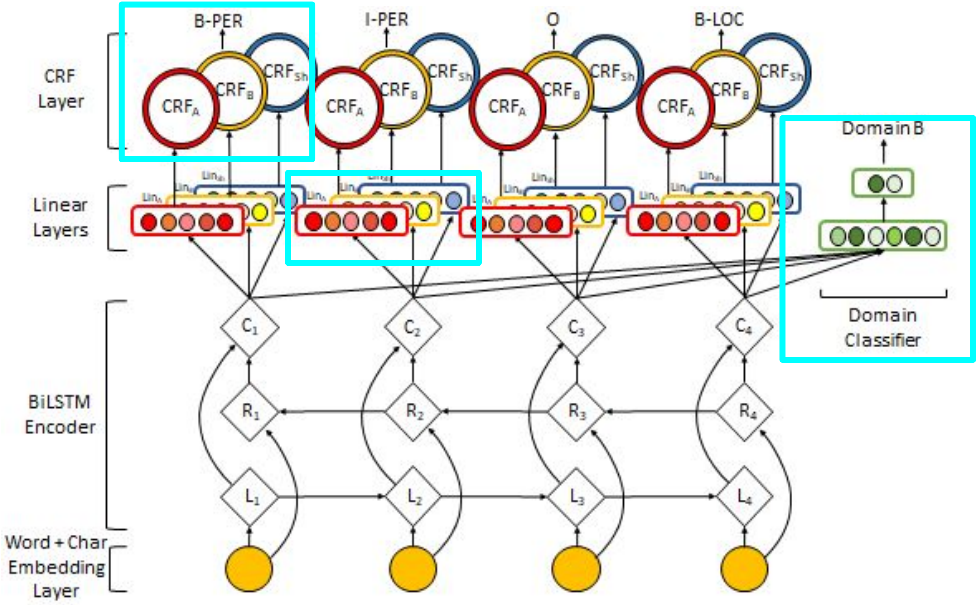
Proposed Model: MultDomain-SP+Aux Domain Learning

Add three components to the base architecture

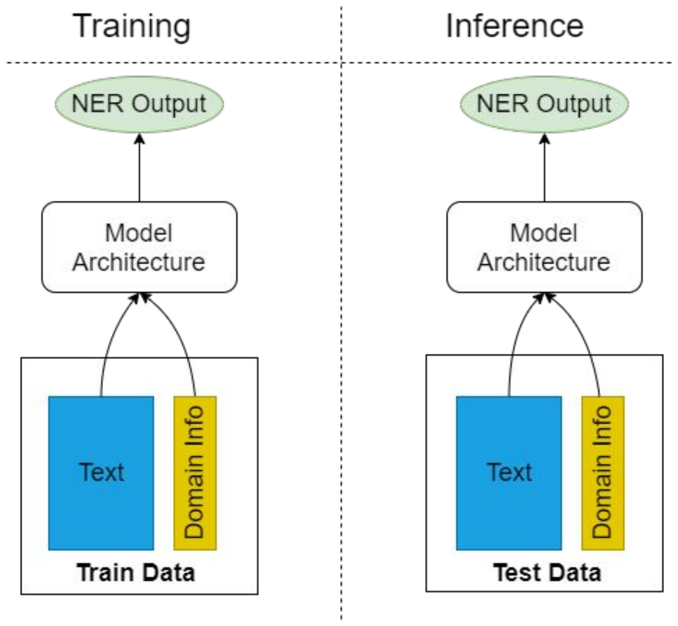
+ domain shared (S) and private (P) CRFs

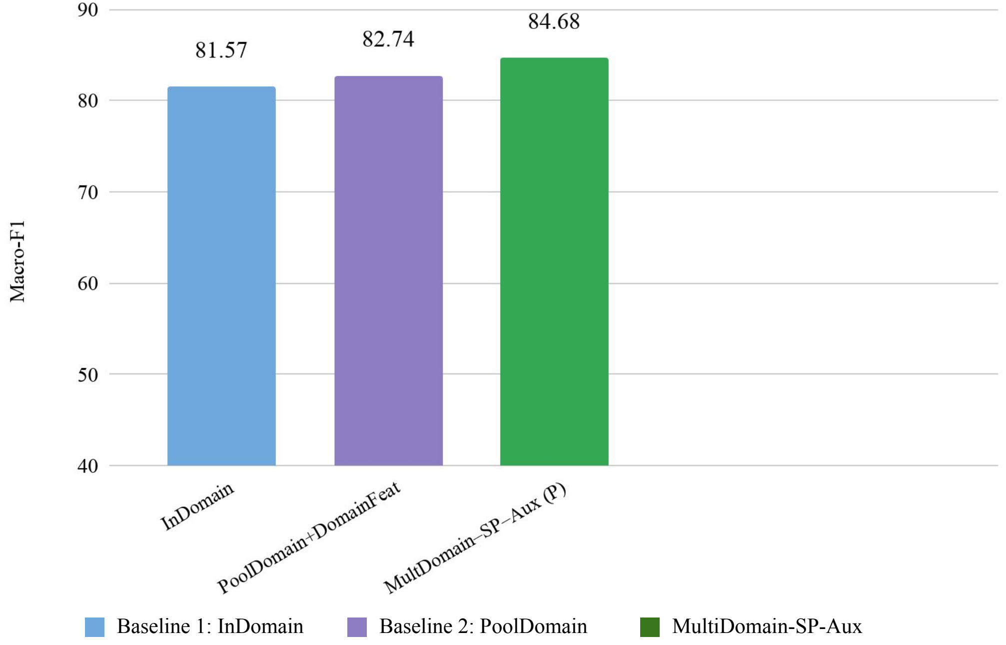
+ domain shared (S) and private (P) linear layers

+ auxiliary task (Aux) of domain classification



Experimental Setup #1: Multiple Domains with specified domain information

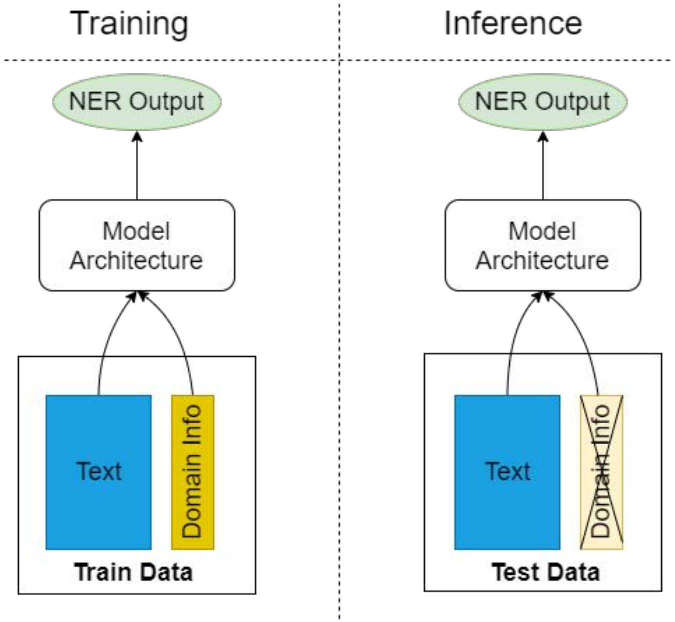


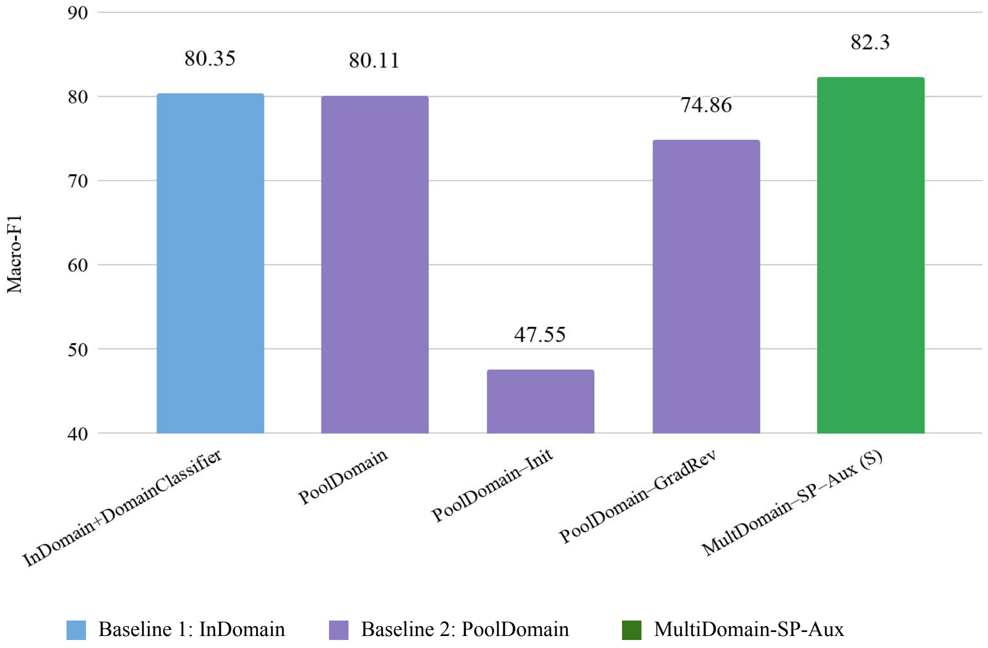


TechAtBloomberg.com

Bloomberg
Engineering

Experimental Setup #2: Multiple Domains without domain information

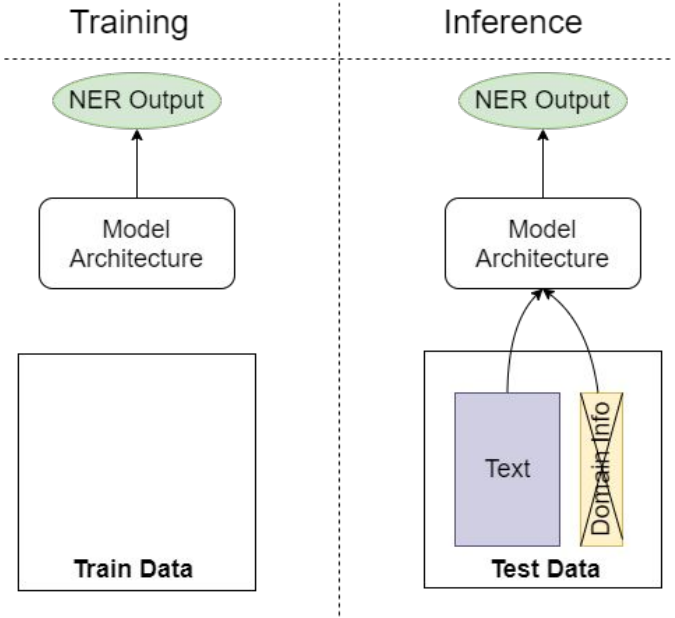


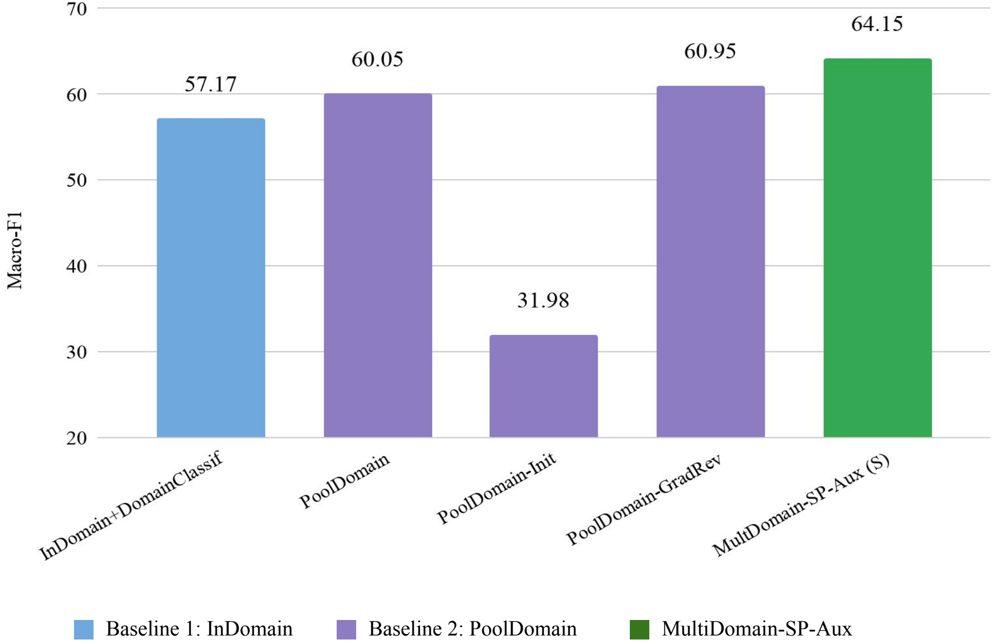


TechAtBloomberg.com

Bloomberg Engineering

Experimental Setup #3: Zero-Shot Domain (Unseen Genres)



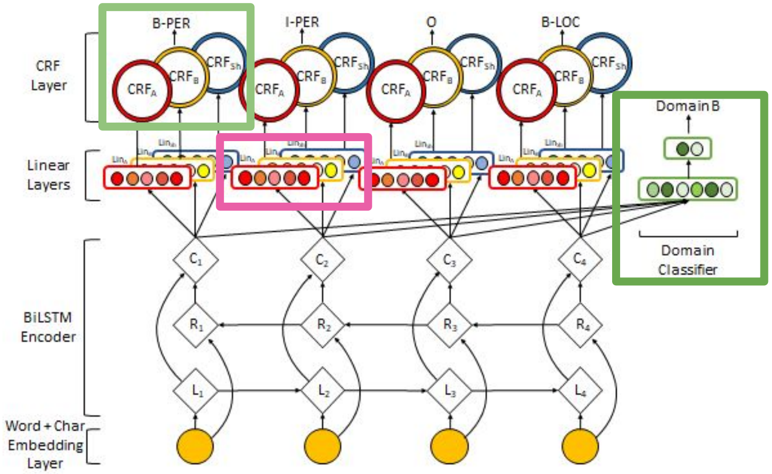


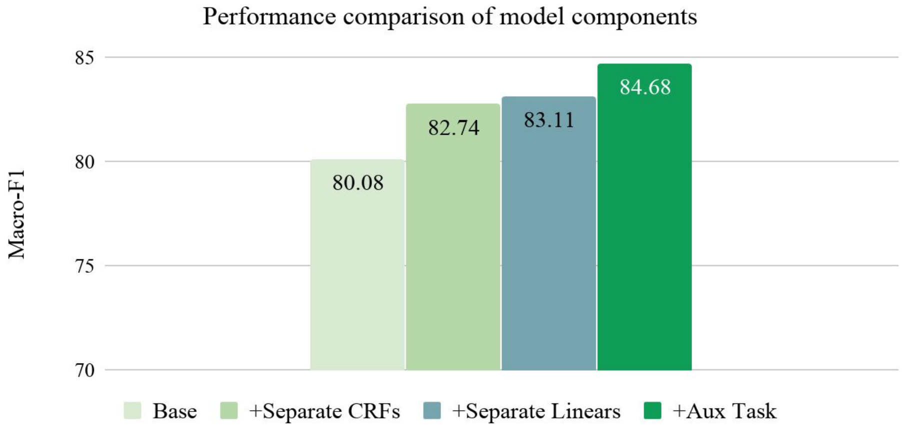
TechAtBloomberg.com

Bloomberg

Engineering

Component Ablation Study





TechAtBloomberg.com

Bloomberg Engineering

Takeaways

- Multi-Domain NER is practically important
- Experiments using three real-world scenarios
- Proposed a robust NER model that works across multiple domains
    - Up to **+5** **F1** compared to baseline approaches

Thank you!

https://www.bloomberg.com/careers

NYC: https://careers.bloomberg.com/job/detail/80630

London: https://careers.bloomberg.com/job/detail/80683

Jing Wang - jwang1621@bloomberg.net
Mayank Kulkarni - mkulkarni24@bloomberg.net
Daniel Preoțiuc-Pietro - dpreotiucpie@bloomberg.net

Q&A Sessions 14B & 15B
Information Extraction 10 & 12
Wednesday, 8 July 18:00 UTC & 21:00 UTC

TechAtBloomberg.com

Bloomberg Engineering

Supplementary Materials

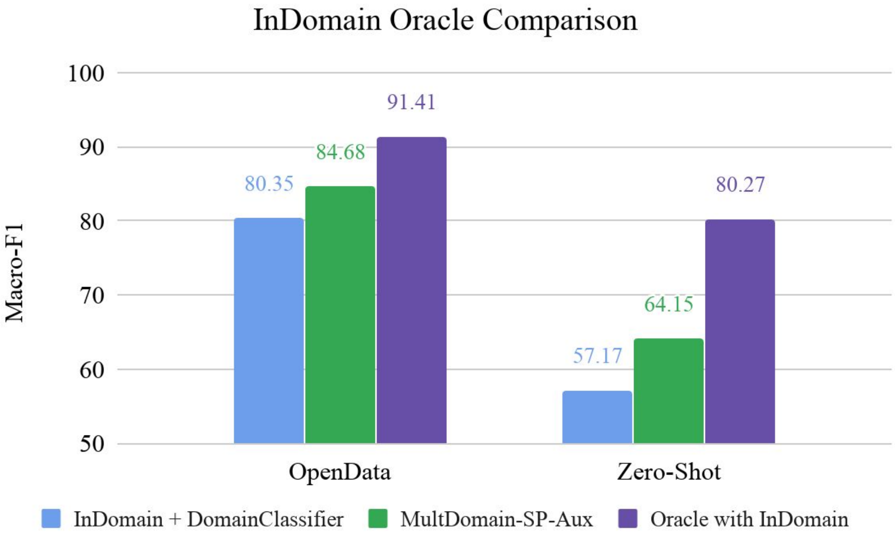
TechAtBloomberg.com

Bloomberg

Engineering

InDomain Oracle Analysis

- Oracle choice across InDomain models by leveraging gold tags
- Upper bound on model performance

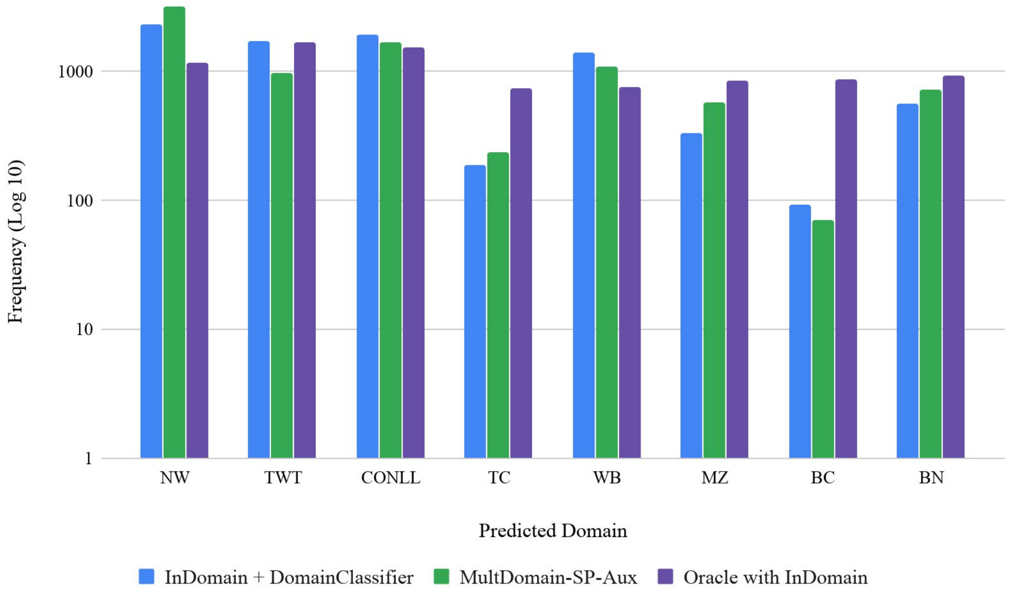


TechAtBloomberg.com

Bloomberg
Engineering

InDomain Oracle - Analysis

- Comparison on domain prediction distribution to analyze where the other models are falling short

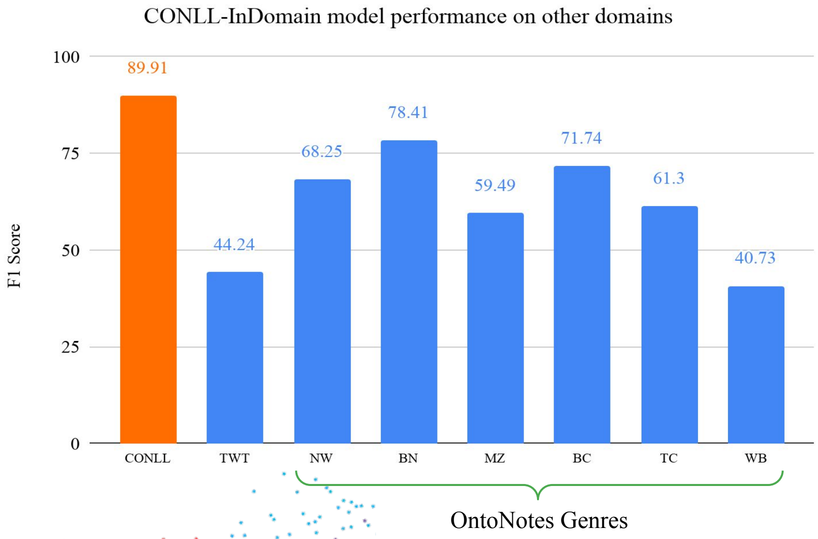


TechAtBloomberg.com

Bloomberg
Engineering

Motivation

- NER models have limited generalization on data from multiple genres
  - High variance of entity mentions
- Existing Setups for NER with multiple domains:
  - Individual bespoke model for each given genre (**InDomain**)
    Overhead of add new genres, Poor generalization



TechAtBloomberg.com

Bloomberg
Engineering

Motivations

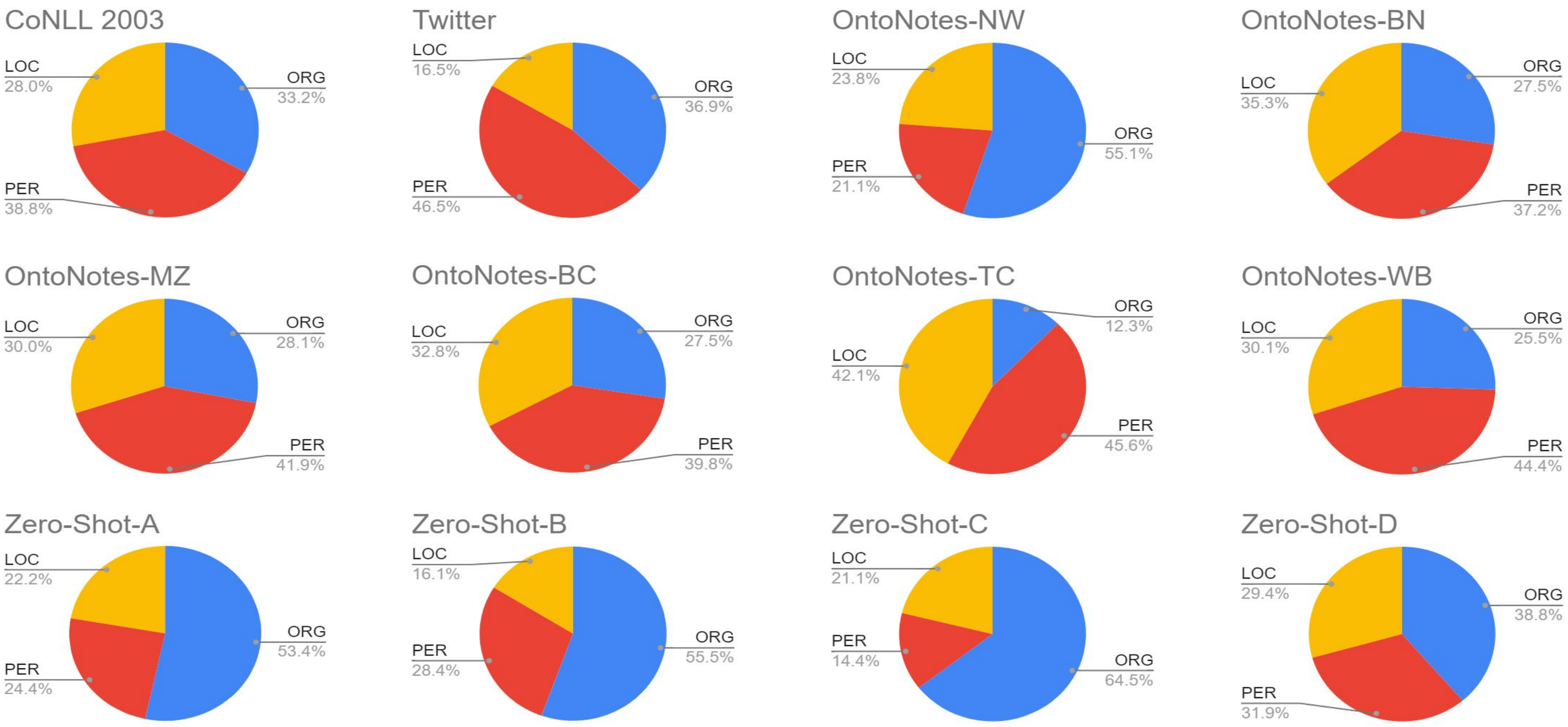
- NER models have limited generalization on data from multiple genres
  - High variance of entity mentions
- Existing Setups for NER with multiple domains:
  - Individual bespoke model for each given genre (**InDomain**)
    Overhead of add new genres, Poor generalization
  - Global model by pooling data from all genres (**PoolDomain**)
    Discard genre-specific information
  - Domain Adaptation between single source domain and single target domain
    Over-simplified scenario for multiple genres
- **MultiDomain**: Jointly learn domain information and NER task from multiple genres of data
  - Leverage genre-specific and genre-independent features
  - Better generalization to unspecified and unseen domains

Key Contributions

- Introduce three experimental setups in inference:
  - Multiple Domains with specified domain information
  - Multiple Domains without domain information
  - Zero-Shot Domain (Unseen Genres)

- Propose a neural architecture for MultiDomain NER with auxiliary domain learning tailored to the above setups

Datasets: Type Distribution Variation across Genres

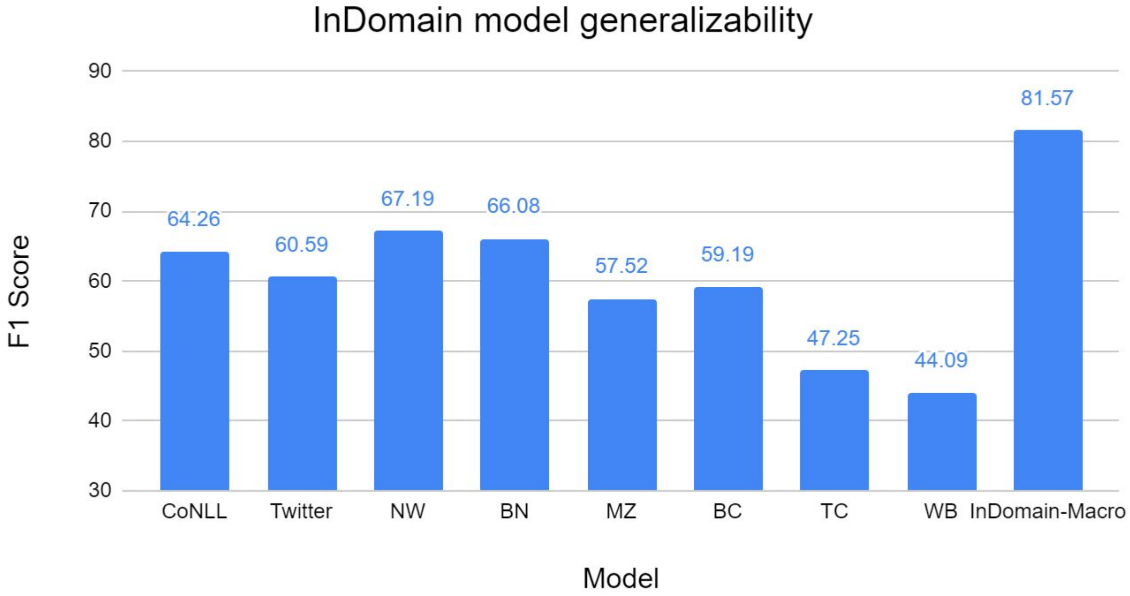


TechAtBloomberg.com

Bloomberg Engineering

InDomain models do not generalize

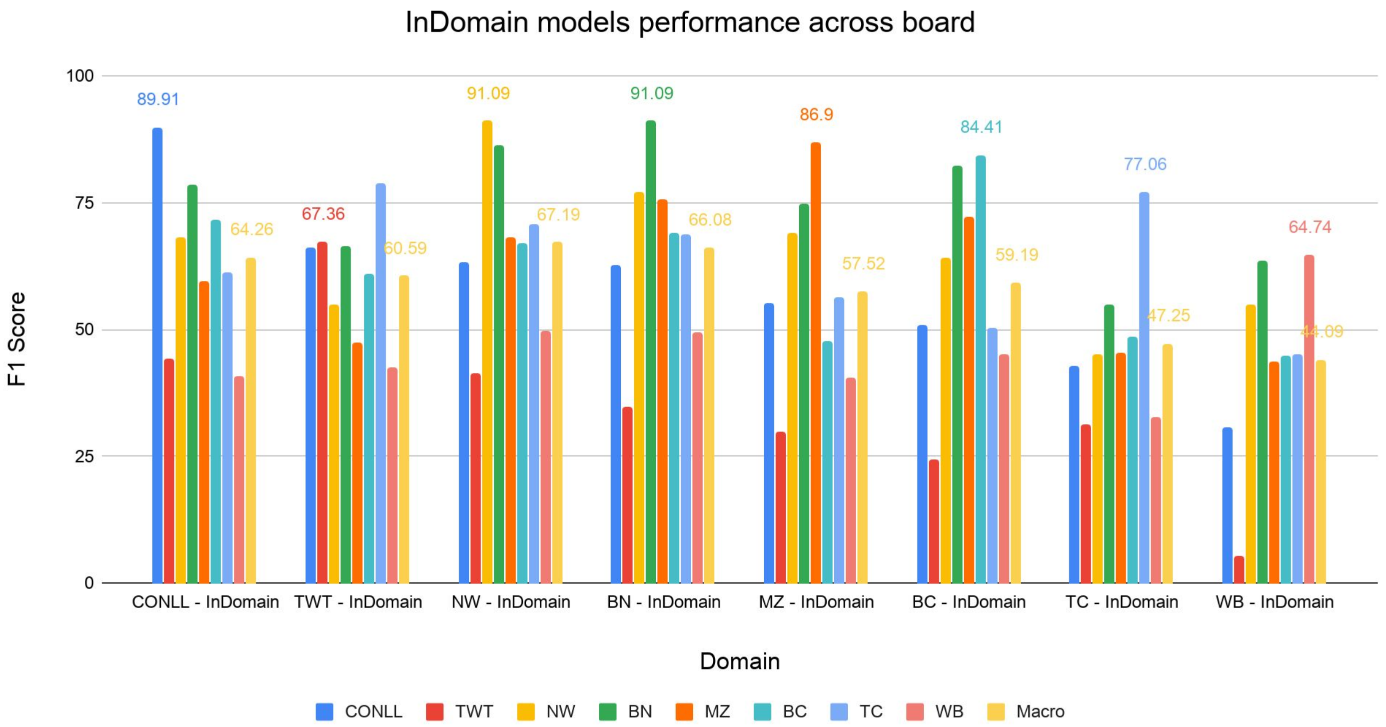
- No single-domain model can generalize across other domains
  - InDomain models for all domains perform significantly worse compared to InDomain models per domain



TechAtBloomberg.com

Bloomberg Engineering

InDomain models performance on rest of the data

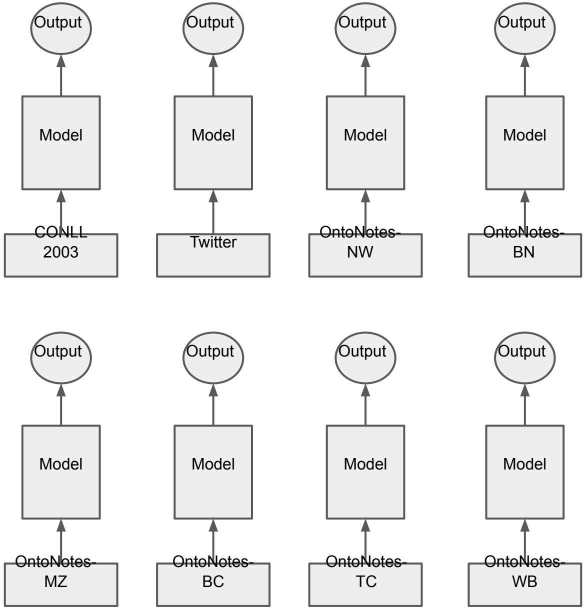


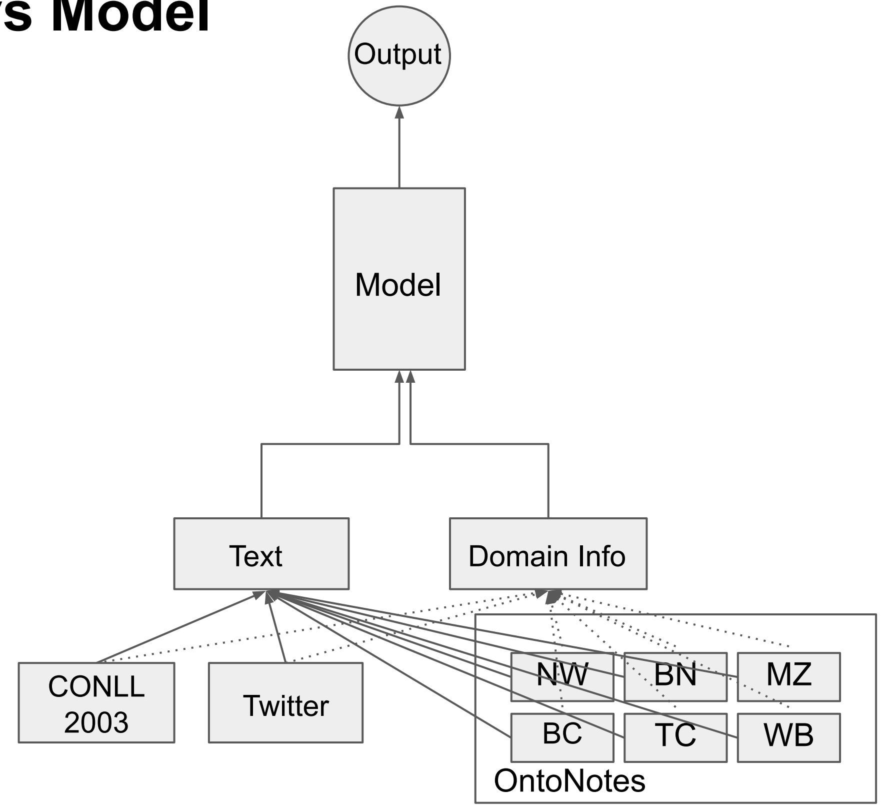
TechAtBloomberg.com

Bloomberg

Engineering

Learning Setup: DataGenre vs Model





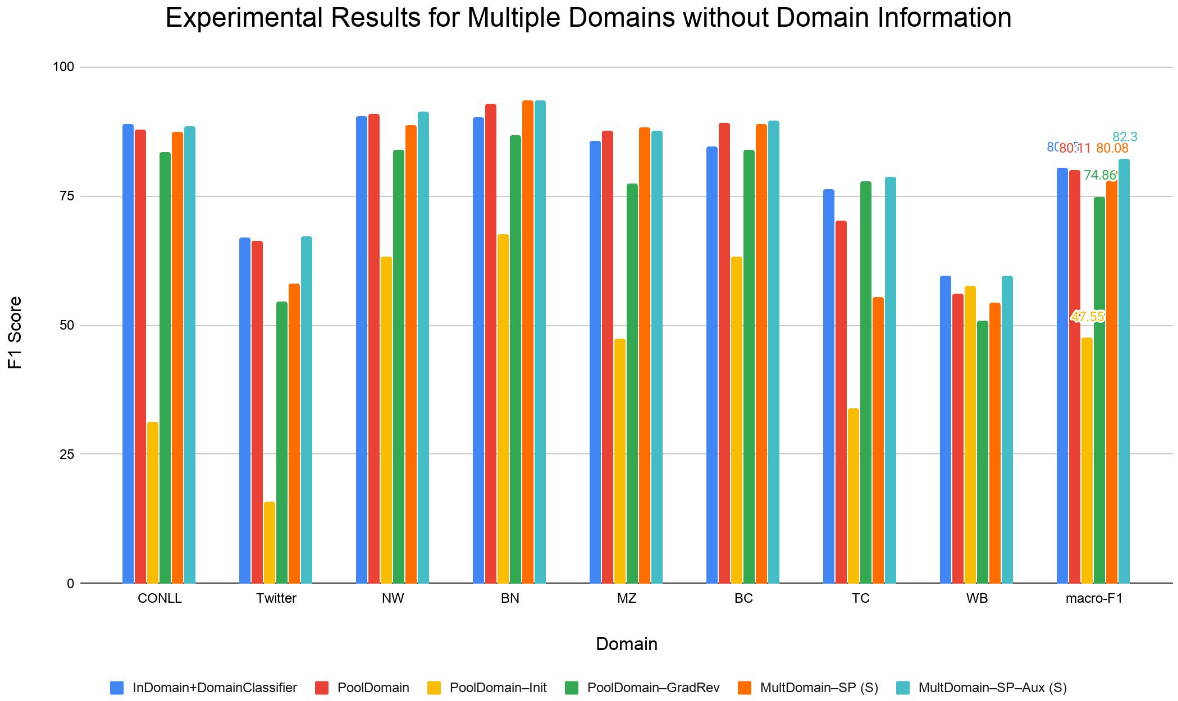
TechAtBloomberg.com

Bloomberg
Engineering

Methods

- DataGenre vs Model: One-to-One
  - **InDomain**
    - InDomain
    - InDomain + DomainClassifier
- DataGenre vs Model: Many-to-One
  - **PoolDomain**
    - PoolDomain
    - PoolDomain-Init
    - PoolDomain-GradRev
    - PoolDomain+DomainFeat
  - **MultDomain-SP**
    - MultCRF-SP
    - MultCRF-SP + Auxiliary Domain Prediction

Results - Multi-Domain without Domain Information

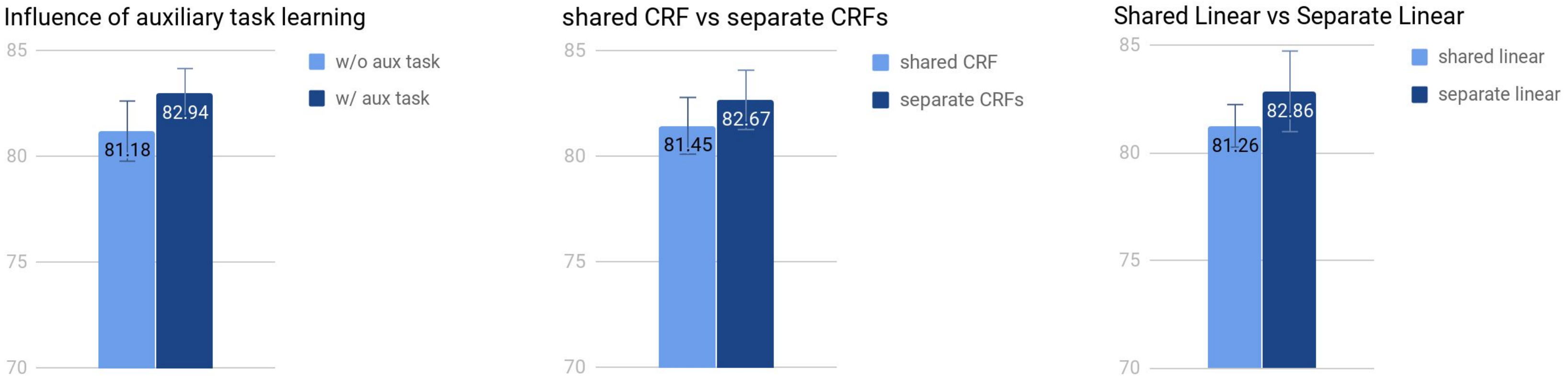


TechAtBloomberg.com

Bloomberg

Engineering

Ablation Studies

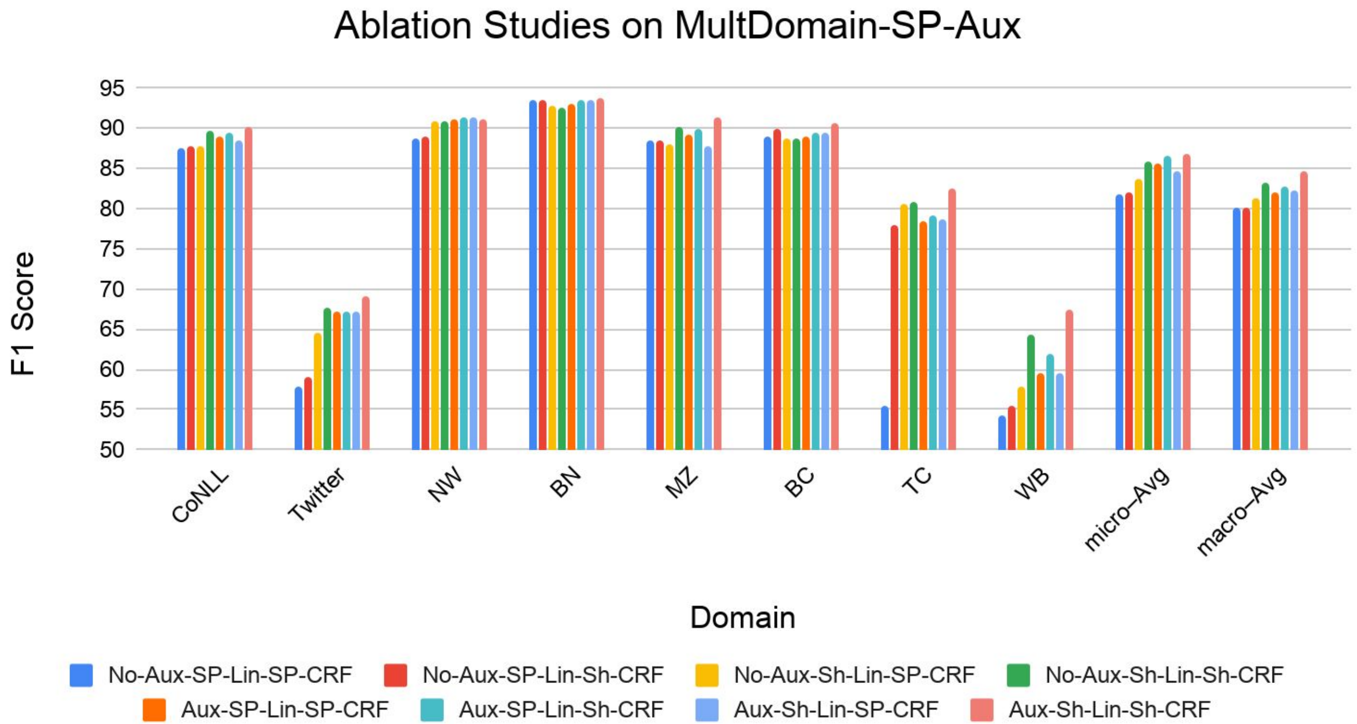


TechAtBloomberg.com

Bloomberg

Engineering

Ablation Studies

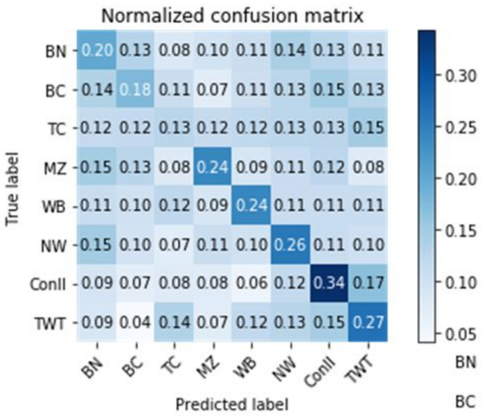


InDomain Oracle - What is it?

- Technique used to calculate upper bound of InDomain model ensembles
- Leverages predictions from InDomain models and gold labels
- Chooses prediction that is closest to gold label based on F1 score
    - If no entities are present, choose model with most ‘O-tag’ predictions
    - If model is tied with multiple other models, select from those models at random

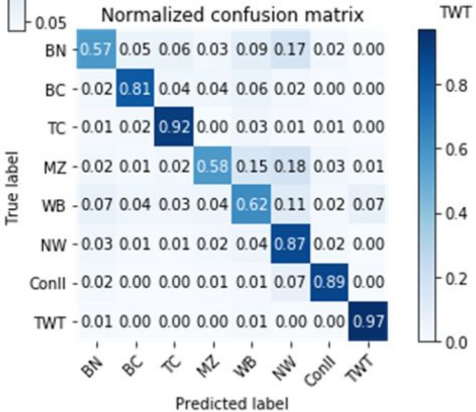
InDomain Oracle - Analysis

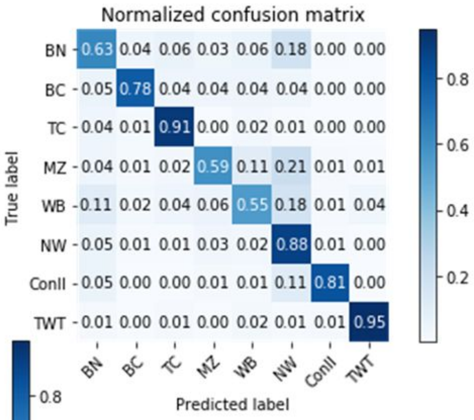
- We compare the domain prediction distribution on Open Data among the Oracle, InDomain+DomainClassifier and MultDomain-SP-Aux, to analyze where the other models are falling short



Oracle with InDomain Models

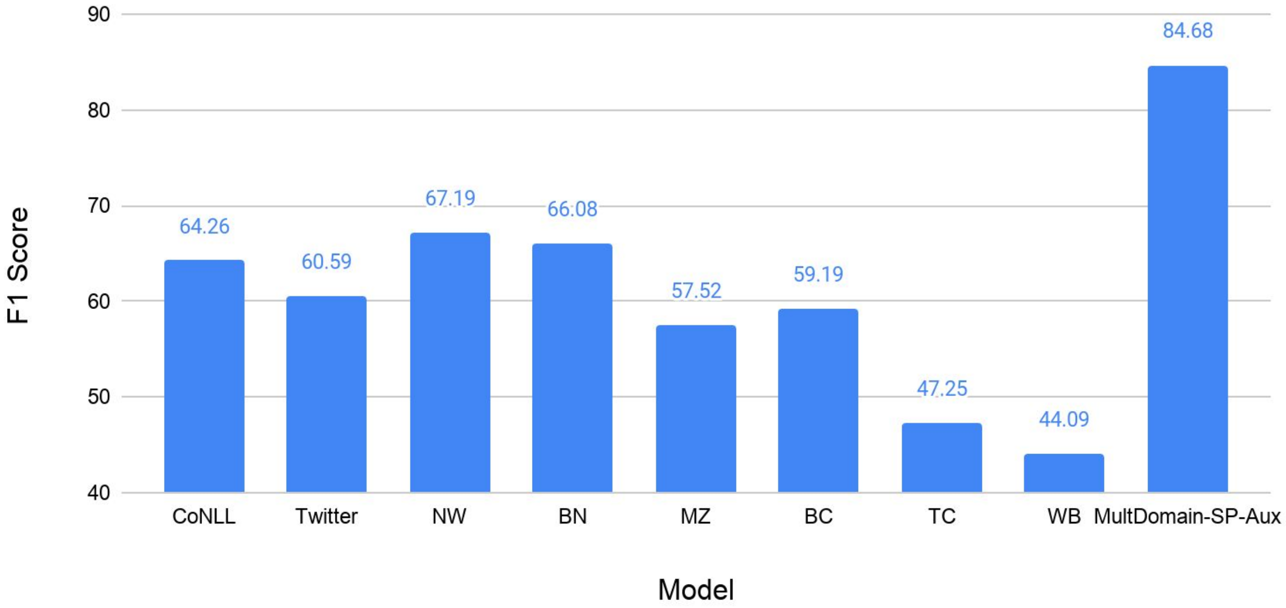
InDomain + DomainClassifier





MultDomain-SP-Aux

Comparison to InDomain Models

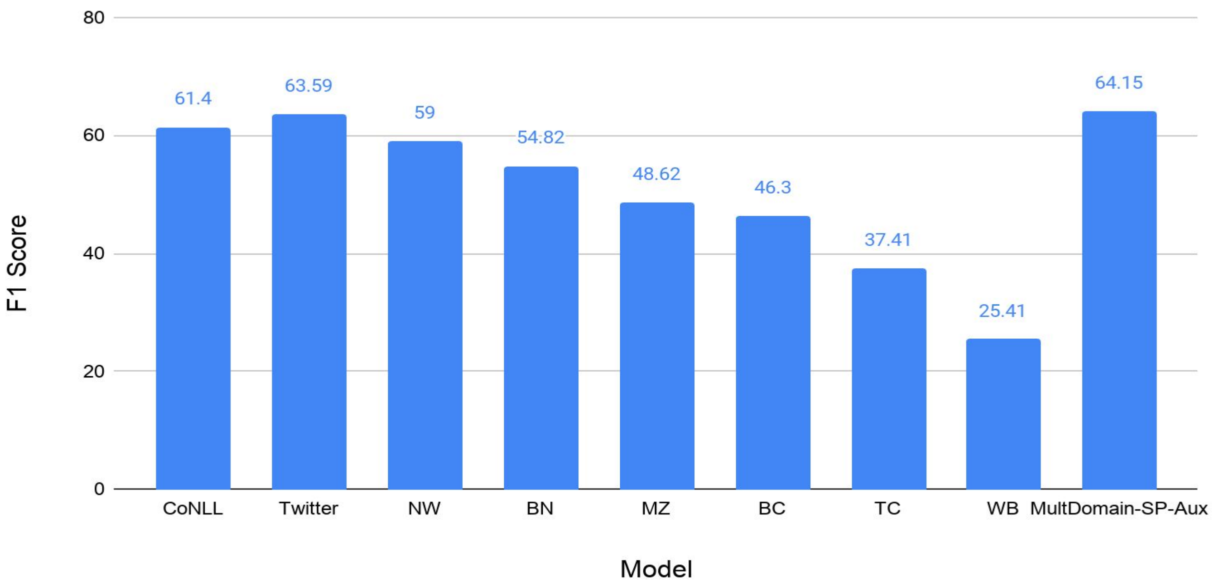


TechAtBloomberg.com

Bloomberg

Engineering

Comparison to InDomain Models on Zero-Shot Genre

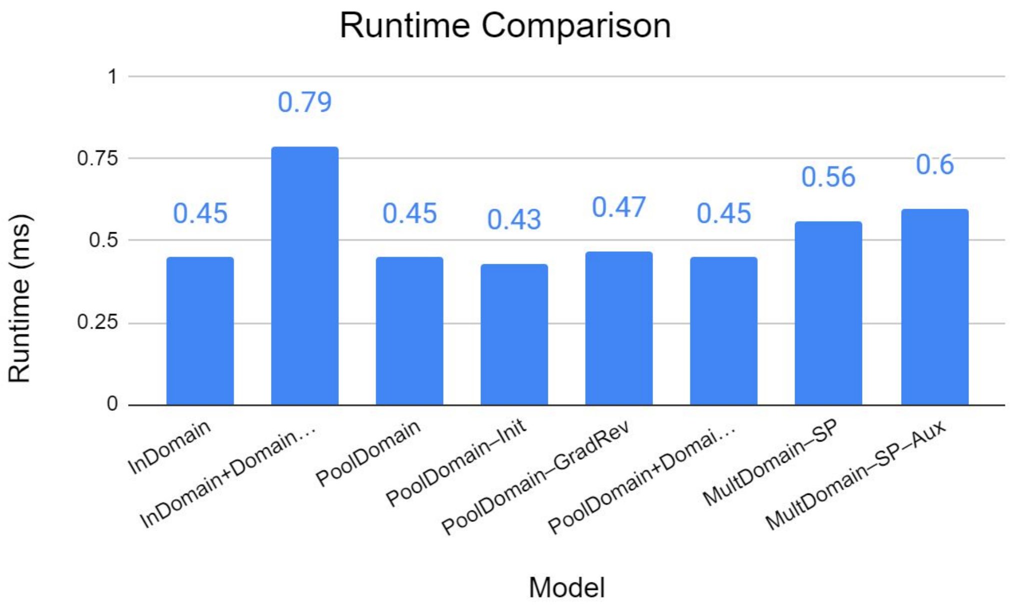


TechAtBloomberg.com

Bloomberg

Engineering

Runtime Comparison



TechAtBloomberg.com

Bloomberg
Engineering