61 ACL 2023

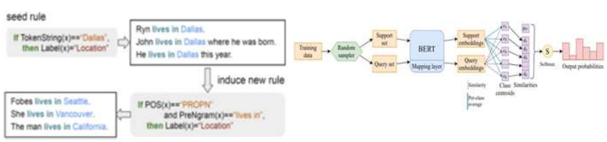
CoAug: Combining Augmentation of Labels and Labelling Rules

Rakesh R Menon, Bingqing Wang, Jun Araki, Zhengyu Zhou, Zhe Feng, Liu Ren

Motivation

- Named Entity Recognition in low resource settings is a challenging problem
- Current approaches:

Rule-based vs Model-based

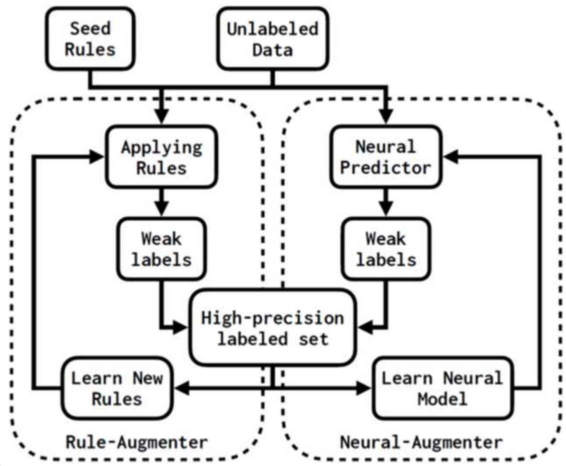


More Accurate More Permissive

Can the complementary skills be combined?

CoAug Framework

- Iteratively train models by bootstrapping augmentations available from both models.



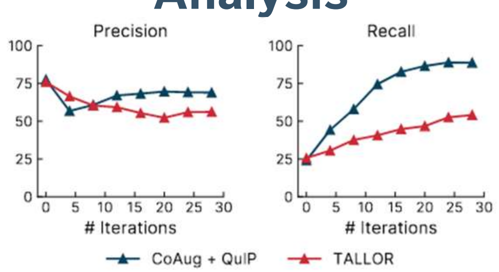
Results

Method	BC5CDR	CoNLL-2003	NCBI-Disease	WikiGold
TaLLOR (Li et al., 2021)	$59.4_{(3.2)}$	$50.3_{(9.6)}$	$39.3_{(1.5)}$	$23.7_{(4.3)}$
ProtoBERT (Tänzer et al., 2022)	$33.1_{(3.5)}$	$47.3_{(2.9)}$	$25.5_{(4.4)}$	$37.3_{(3.8)}$
CoAug (TaLLOR + ProtoBERT)	$\mathbf{64.4_{(1.5)}}$	$\mathbf{65.0_{(0.8)}}$	$\mathbf{46.8_{(3.5)}}$	$\mathbf{50.6_{(2.1)}}$
QuIP (Jia et al., 2022)	$64.9_{(1.7)}$	$70.6_{(3.7)}$	$\mathbf{75.3_{(0.7)}}$	$43.6_{(2.3)}$
CoAug (TaLLOR + QuIP)	$\mathbf{65.9_{(1.5)}}$	$\mathbf{76.8_{(2.0)}}$	$50.5_{(4.9)}$	$\mathbf{51.8_{(2.8)}}$

F1 on Test Set for four datasets spanning science-domain and general domain

- CoAug (TaLLOR + ProtoBERT) is better than individual components across datasets ≥ 9.5 F1 points.
- CoAug (TaLLOR + QuIP) is better than QuIP on three datasets.
- On NCBI, noisy weak labels in early iterations result in lower F1 (still higher than TaLLOR).

Analysis



Performance improvement for CoAug comes from improved recall while maintaining high-precision

Conclusion

- We present an iterative framework that combines the best of neural and rule-based models for low-resource NER.
- Future work will focus at improving robust-ness and efficacy of the framework

UNC NLP
Contact: rrmenon@cs.unc.edu
Code: https://github.com/boschresearch/CoAug