Transfer Learning for Entity Recognition of Novel Classes

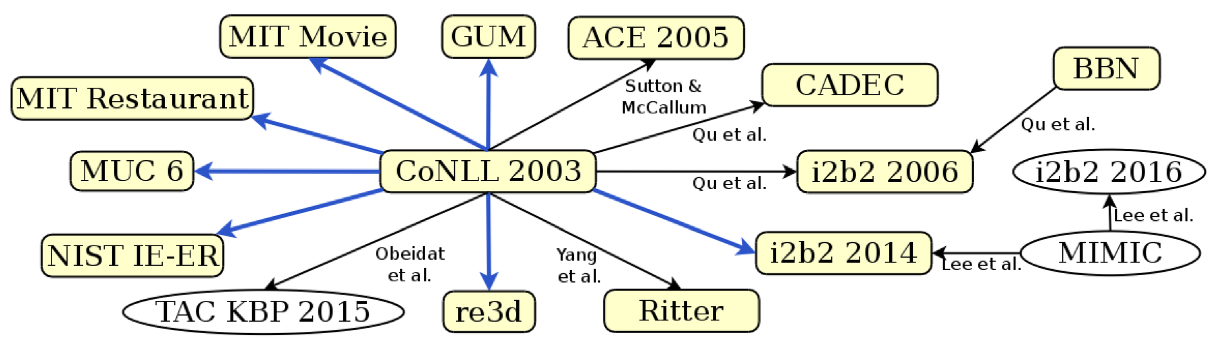
Juan Diego Rodriguez, Adam Caldwell and Alexander Liu

Applied Research Laboratories, The University of Texas at Austin

Objectives

We replicated several studies on transfer learning for entity recognition where the class labels in the source and target domains are different.

1. Compared transfer learning methods for entity recognition that had not been compared before.
2. Evaluated across a range of target training set sizes.
3. Performed experiments on seven new source/target corpus pairs.
4. Code available at: `http://github.com/ciads-ut/transfer-learning-ner`



Source/target domain pairs

Transfer learning methods

Neural Methods

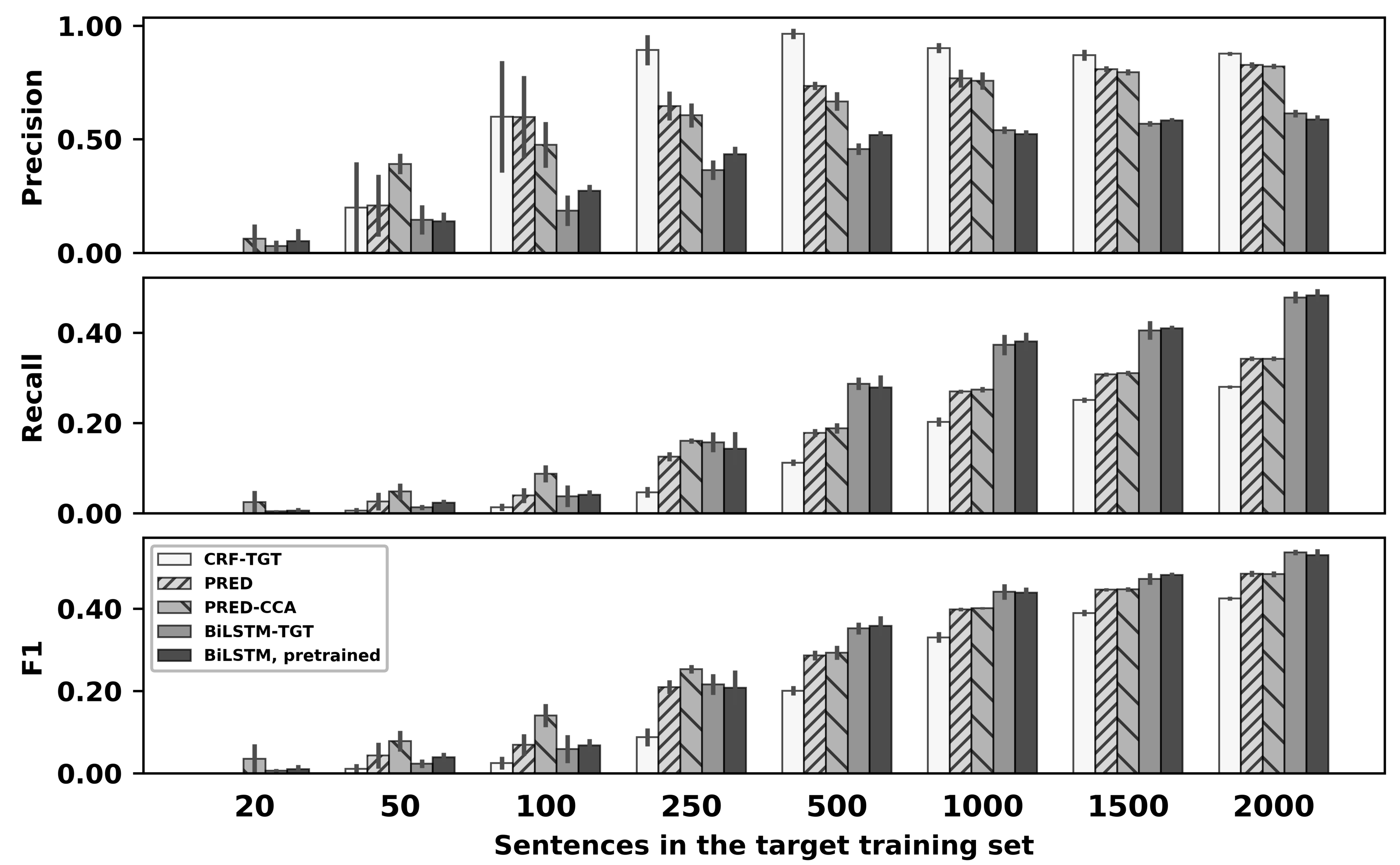
- **BiLSTM-CRF**: a bidirectional LSTM-CRF trained on the target domain (no transfer).
- **BiLSTM-CRF, pretrained ([2])**: a bidirectional LSTM-CRF trained on the source domain, and fine-tuned on the target domain.

Non-Neural Methods

- **CRF (Conditional Random Field)**: a linear CRF trained on the target domain (no transfer).
- **PRED:** the outputs of a CRF trained on the source domain are used as features for a second CRF trained on the target domain ([1]).
- **PRED-CCA ([3]):** a variant of PRED, where:
  - both the source and target labels are first embedded in a common space using CCA (Canonical Correlation Analysis).
  - the second CRF uses the label embeddings as features.

Results

Micro-averaged scores for the Ritter corpus



Datasets

Corpus	Domain	Annotated entities
Ritter	Twitter	*Person*, *geo-loc*, facility, company, sportsteam, band, product, tv-show, movie, other
i2b2 2006	Medical	Date, doctor, hospital, ID, *location*, patient, phone
i2b2 2014	Medical	Age, username, ID, patient, doctor, profession, hospital, street, state, zip, city, country, *organization*, date, medical record, phone
CADEC	Medical	Adverse reaction, disease, drug, finding, symptom
MUC 6	News	*Person*, *location*, *organization*, date, percent, money
NIST IE-ER	News	*Person*, *location*, *organization*, date, duration, percent, money, cardinal
GUM	Wikinews wikihow wikivoyage	Abstract, animal, event, object, *organization*, *person*, *place*, plant, quantity, substance, time
MIT Movie	Spoken queries	Actor, character, director, genre, plot, year, soundtrack, opinion, award, origin, quote, relationship
MIT Restaurant	Spoken queries	Amenity, cuisine, dish, hours, *location*, price, rating, restaurant name
re3d	Defense	Document reference, *location*, military platform, money, nationality, *organization*, *person*, quantity, temporal, weapon

Summary and Conclusions

With enough data, neural approaches outperform non-neural approaches due to increased recall.

- However, PRED/PRED-CCA had greater precision than the neural network approaches in almost every case.
- The neural network approaches had greater recall than PRED/PRED-CCA for the MIT Restaurant, MIT Movie, CADEC, i2b2 2014 and i2b2 2006 datasets, across all dataset sizes. For the other datasets, the neural approaches had greater recall than the non-neural approaches only when the target training dataset was large enough.

Generally, CRF-TGT < PRED < PRED-CCA.

- For most corpora, the PRED/PRED-CCA approaches improved the F1 score over the target-only CRF baseline.
- PRED-CCA outperformed PRED in most cases. The improvement was due largely to an increase in recall, and was greatest for smaller datasets.

Pre-training the BiLSTM-CRF generally outperforms training the neural network from scratch, but the difference is small.

- On most of the datasets, using pre-training was superior to training the BiLSTM-CRF from scratch.
- This mainly occurred when there were 500 target training sentences or less. The improvement was often between 2% and 3%. However, in most of these cases, pre-training was outperformed by PRED-CCA.

Future Research

- Experiments with different source domains
- Adapt few-shot learning approaches from image classification
- Error analysis based on the relation between source and target labels

References

[1] Hal Daumé III. 2007. Frustratingly easy domain adaptation. *In Proceedings of ACL 2007.*

[2] Ji Young Lee, Franck Dernoncourt, and Peter Szolovits. 2017. Transfer Learning for Named-Entity Recognition with Neural Networks. *arXiv preprint arXiv:1705.06273.*

[3] Rasha Obeidat, Xiaoli Z. Fern, and Prasad Tadepalli. 2016. Label embedding approach for transfer learning. *In Proceedings of the Joint International Conference on Biological Ontology and BioCreative, 2016.*