NER Applications in Topic Identification and Few-Shot Learning

Alex Feng[2], Hope McGovern[1], Richard Sear[3], Simon Zeng[4], Francis Ferraro[5], Clare Grasso[5]

[1]Brown University, [2]The College of William and Mary, [3]The George Washington University, [4]Johns Hopkins University, [5]The University of Maryland, Baltimore County

We investigate both standard and fine-tuned performance over sparsely-labeled datasets and examine the possibility of transfer learning between NER tasks and topic identification tasks.

Varying Entity Counts

Here, we introduce a framework for varying the number of instances of particular entity labels in a training set. This is useful for simulating a sparsely-labeled environment or the introduction of a new entity type.

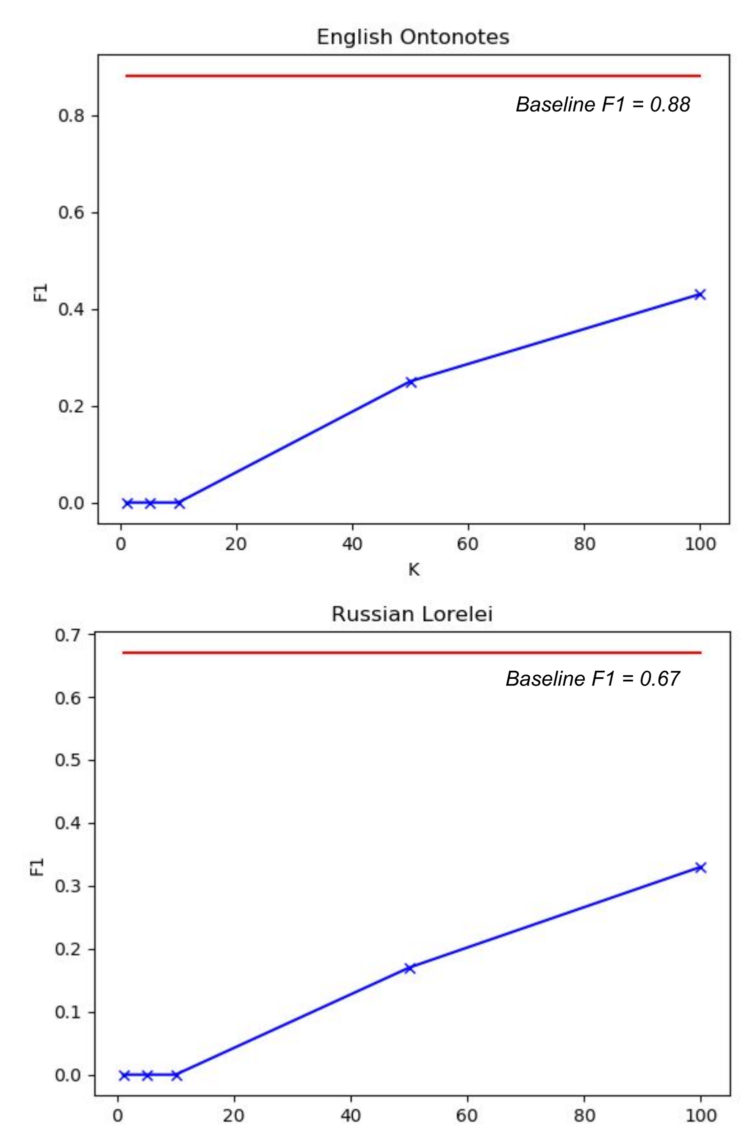

Figure 1. NER results with fastText embeddings with all entity types restricted

"K" refers to the number of each entity type in the training set. For example, a model trained with K=1 only saw 1 'PER', 1 'ORG', 1 'LOC' and 1 'GPE'.

Results show little difference at low values of K. The steady increase afterwards suggests some "threshold" value at which models begin to gain traction. We then conduct experiments with only one restricted entity type to loosely simulate the introduction of a new entity.

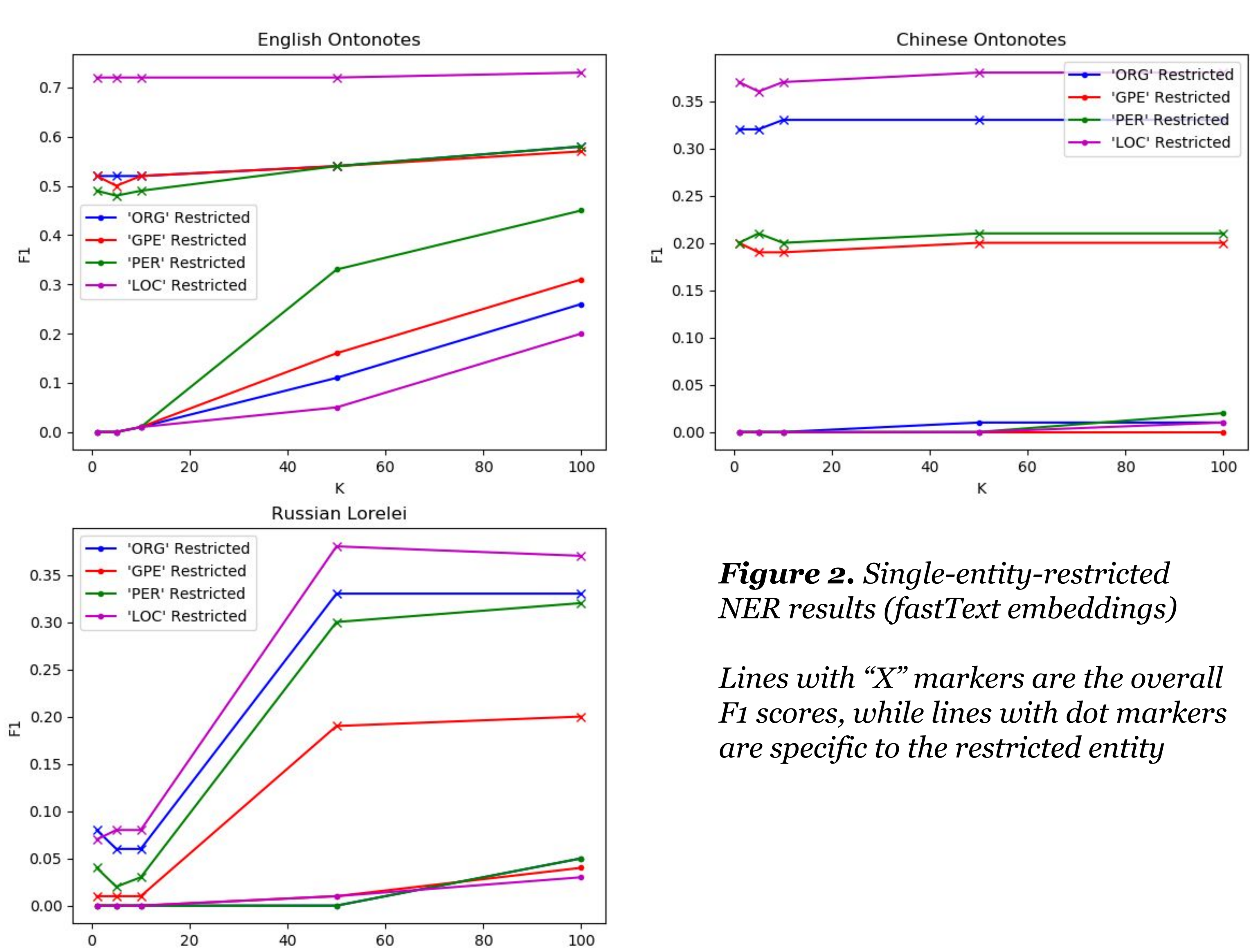

Figure 2. Single-entity-restricted NER results (fastText embeddings)

Lines with "X" markers are the overall F1 scores, while lines with dot markers are specific to the restricted entity

Iterative Fine-Tuning

Fine-tuning a model on one task may improve performance on subsequent tasks. To investigate this, we set up the following framework for iteratively fine-tuning:

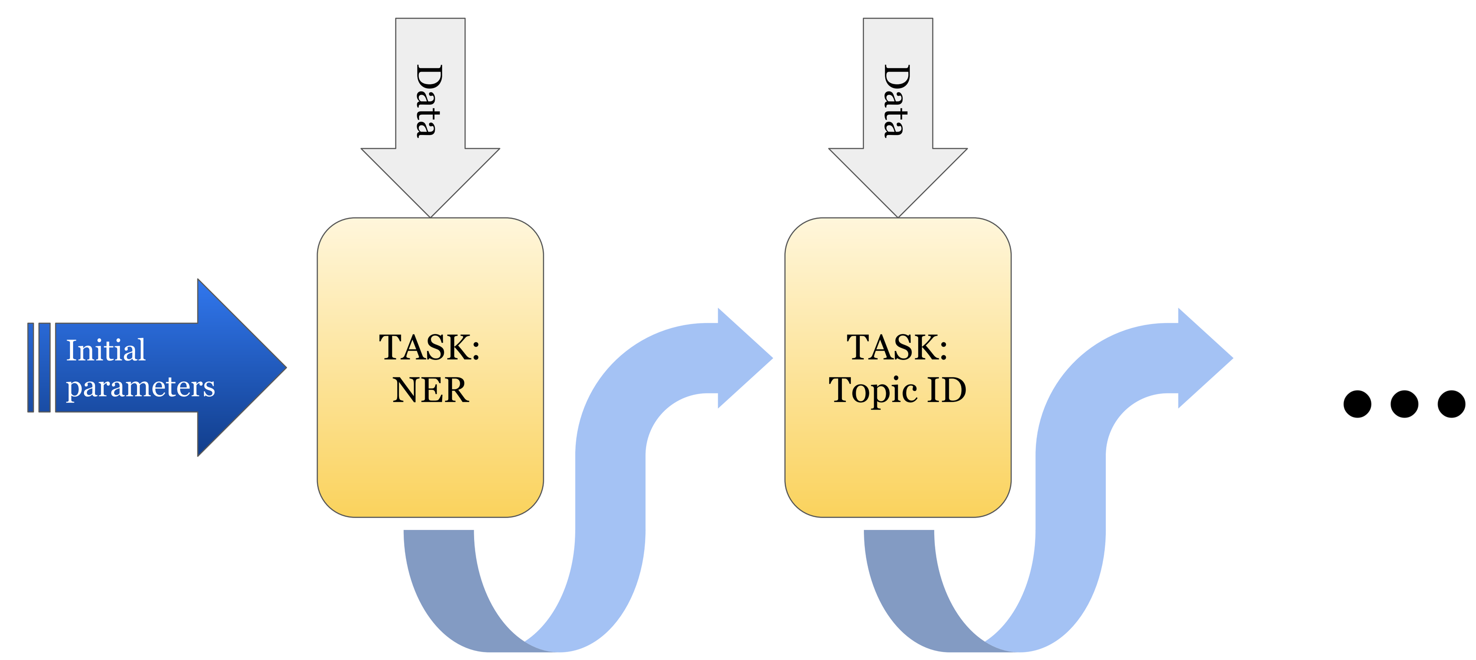

Figure 3. *Visual depiction of the iterative fine-tuning framework*

Here, we fine-tune BERT on an NER task with Wikipedia data and a Topic ID task with DLI data, and evaluate on an NER task with Chinese Ontonotes.

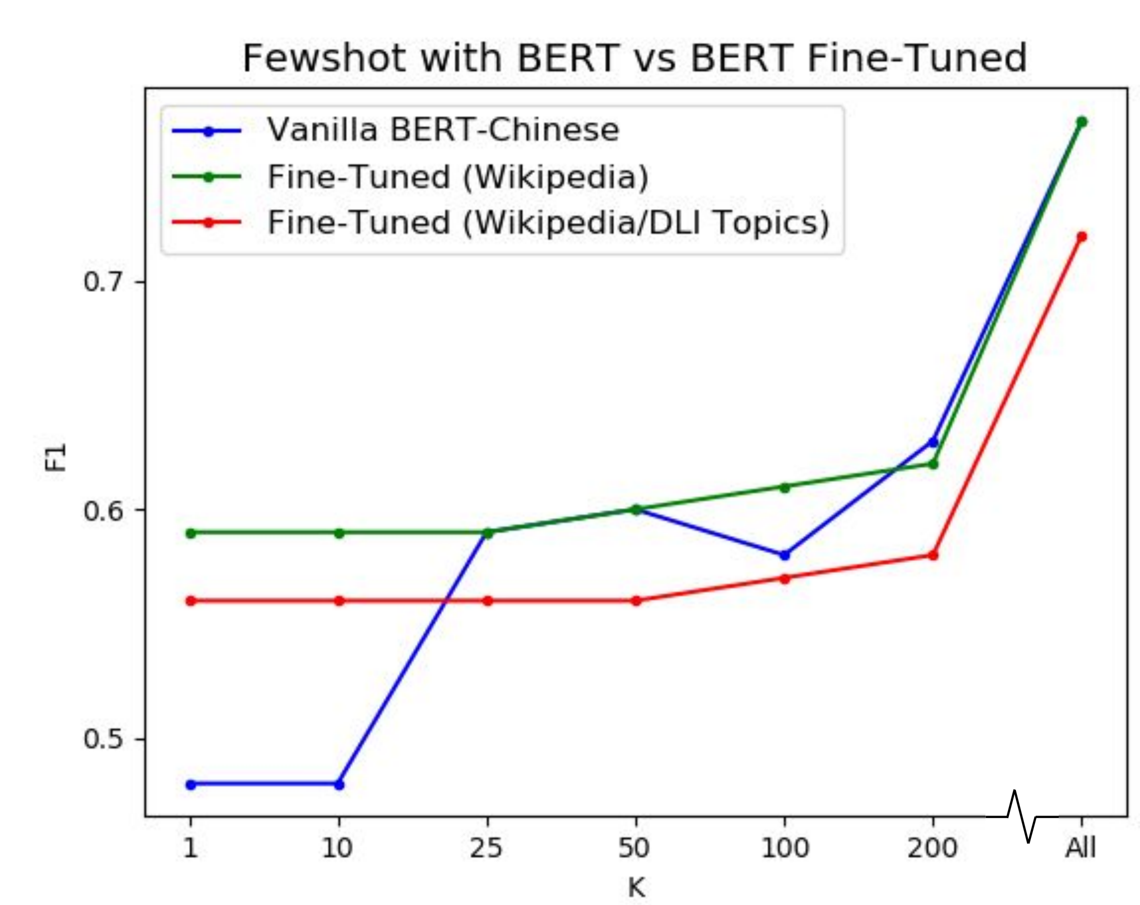

Figure 4. *BERT fine-tuning experiments with K = 1, 10, 25, 50, 100, 200*

Multi-Label Topic Identification

Separately, we establish a framework for multi-label topic identification for given documents. The datasets used display significant imbalance between topic classes, as shown in **Figure 5**.

	Culture	Society	Defense	Security	Ecology	Economics	Geography	Technology	Science	Politics
dli_ru	439	855	206	208	43	533	150	207	176	454
nflc_ru	39	39	7	7	11	36	11	17	17	36
dli_cn_all	1474	1543	238	238	103	1310	432	453	365	1114
nflc_cn	166	166	51	51	56	75	56	49	49	75

Figure 5: *Document Topic Label Distribution of Topic ID Datasets*

We analyze the effects of using different document embeddings (fastText and BERT) as well as non-neural and neural architectures:

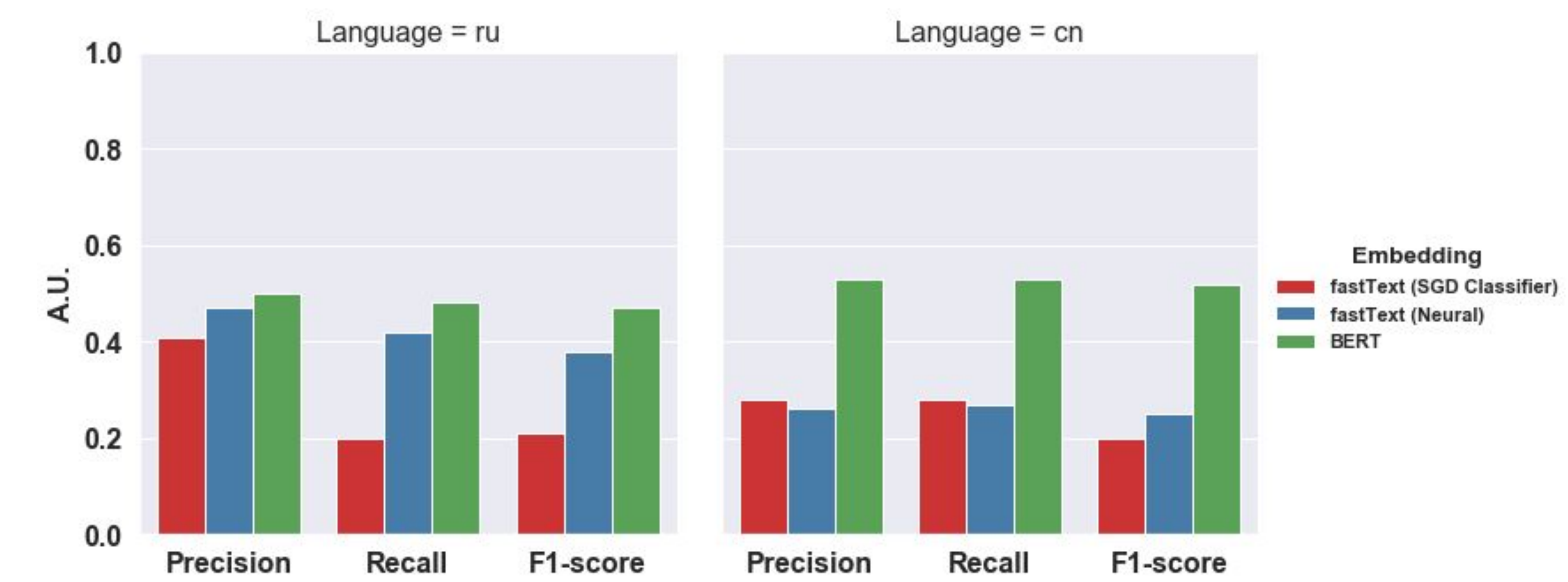

Figure 6: *Comparison of results on the DLI datasets between a non-neural classifier (SGD) using fastText embeddings and two neural classifiers, one with fastText embeddings and the other with BERT embeddings. Dropout of 0.3.*

As seen in **Figure 6**, the BERT models far outperform either model with fastText word embeddings, revealing that, for our data, the embedding method is a more influential factor in determining performance than the model architecture itself.

Additionally, we fine-tune BERT on a separate NER task (Russian Reddit for Russian and Chinese Ontonotes for Chinese) before then re-running the Topic ID experiment.

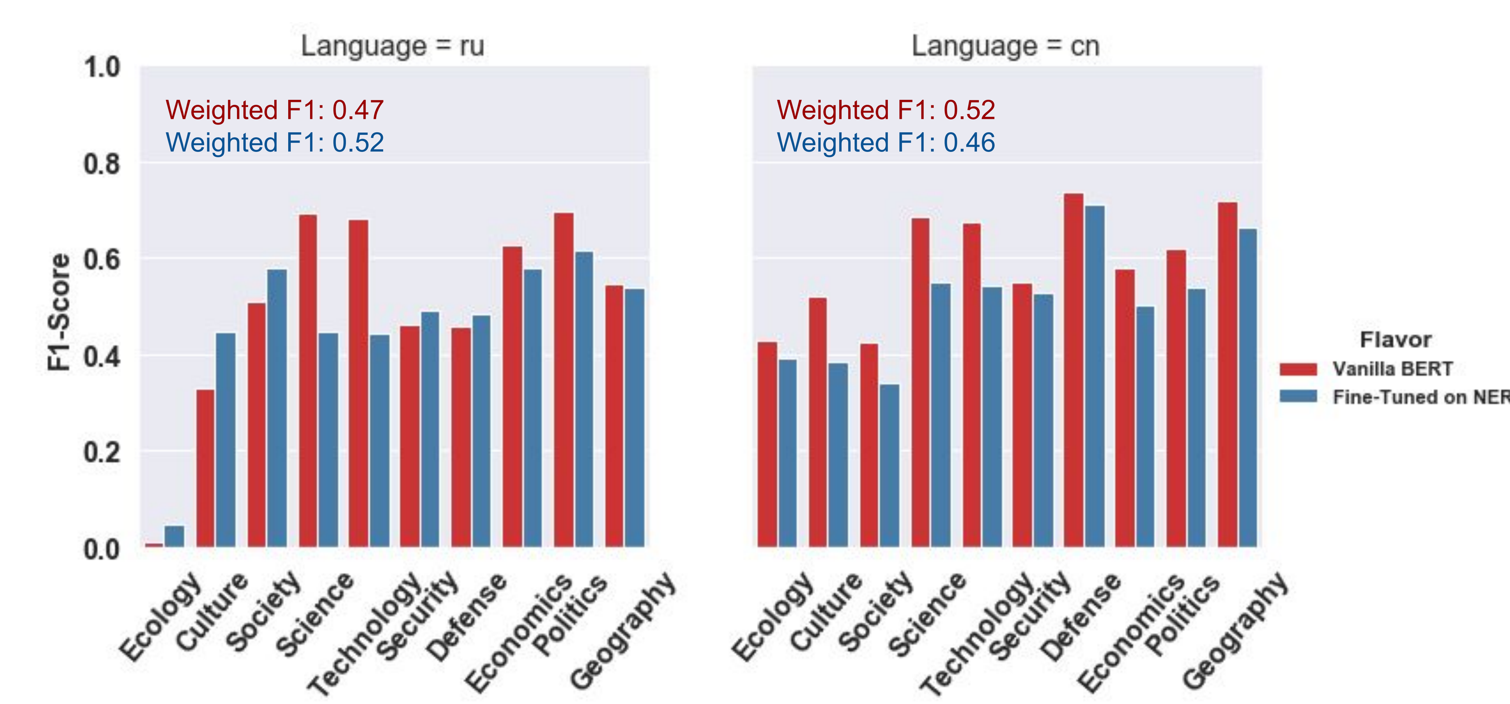

Figure 7: *Comparison between vanilla BERT and fine-tuned BERT on Topic ID with the DLI datasets. All using the same multilayer perceptron model structure and hyperparameters.*

The results of the two-step BERT fine-tuning process display an increase in the weighted F1-score when evaluating on the Russian DLI dataset, due to increased performance in the two overrepresented categories, 'Culture' and 'Society'. However, the fine-tuning decreased weighted F1-scores for the Chinese DLI dataset. This discrepancy may have to do with structural differences in the Russian Reddit and Chinese Ontonotes datasets which were used for the first step of fine-tuning, respectively.

Acknowledgements

This research would not have been possible without the sage guidance of Dr.'s Francis Ferraro and Clare Grasso, the wonderful SCALE leads (Dr.'s Nicholas Andrews, Dawn J. Lawrie, and James Mayfield), the revitalizing energy of SCALEdown, and finally, our fellow student participants.