Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee and Luke Zettlemoyer. Deep contextualized word representations. Proceedings of NAACL-HLT 2018, pages 2227–2237.
https://aclweb.org/anthology/N18-1202

Deep Contextualized word representations (**E**mbeddings for **L**anguage **Mo**dels) for Named Entity Recognition

Input/s:

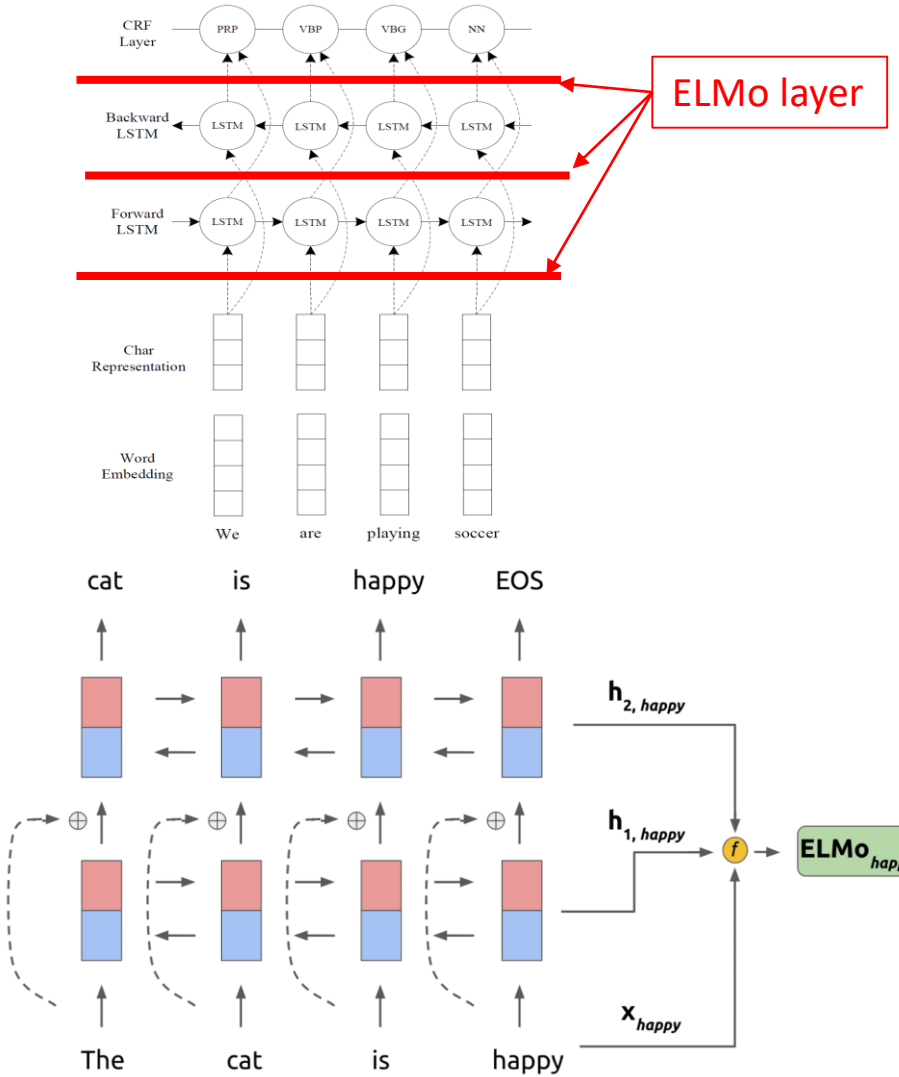
- Text as sentence or tokenized strings
- Glove – glove.6B.50d

Output/s:

- BIO tagged Entities

Approach:

- Tokenize the input and fetch the glove representations for each token
- Feed these tokenized sentences in the reverse order into model. This paper uses a stacked, multi-layer(L) LSTM. Whereas a single-layer LSTM would take the sequence of words as input, a multi-layer LSTM trains multiple LSTMs to take the output sequence of the LSTM in the previous layer as input
- Add the original word vectors, we have 2L + 1 vectors that can be used to compute the context representation of every word
- Using these word representations the CRF layer estimates the best tag candidate with highest probability.



https://github.com/blackbbc/NER

Sarvesh Parab

Evaluation Datasets:

Named Entity Recognition Datasets

- CoNLL ’03:
  - English Dataset contains collection of news wire articles from Reuters
  - Per-Loc-Org-Misc tagging
  - *Dataset used in the paper as well.*
- OntoNotes 5.0:
  - Includes text from various genres of text like news, weblogs, broadcast, newsgroups etc.
  - 15+ tags

Evaluation metrics:

- Precision
- Recall
- F1 Score

Model	$F_1 \pm$ std.
Collobert et al. (2011)♣	89.59
Lample et al. (2016)	90.94
Ma and Hovy (2016)	91.2
Chiu and Nichols (2016)♣,♢	91.62 ± 0.33
Peters et al. (2017)♢	91.93 ± 0.19
biLSTM-CRF + ELMo	**92.22 ± 0.10**

```
  "Converting sparse IndexedSlices to a dense Tensor of un
[DeepElmoEmbedNer.py:138 - train() ] Start training...
[DeepElmoEmbedNer.py:139 - train() ] Train size = 14041
[DeepElmoEmbedNer.py:140 - train() ] Val size = 3453
[DeepElmoEmbedNer.py:141 - train() ] Test size = 3250
[DeepElmoEmbedNer.py:142 - train() ] Num classes = 9
```

```
04-23 21:37 epoch: 96, size: 14032/14041, step_loss: 2.951782, epoch_loss: 32807.906706
04-23 21:37 epoch: 96, size: 14041/14041, step_loss: 2.336548, epoch_loss: 32810.243254
04-23 22:13 Epoch: 96, val_f1: 90.790246
04-23 22:50 Epoch: 96, test_f1: 87.802730
04-23 22:50 checkpoints/best\best.ckpt-43 is not in all_model_checkpoint_paths. Manually adding it.
04-23 22:50
04-23 22:50 epoch: 97, size: 16/14041, step_loss: 17.713564, epoch_loss: 17.713564
```

Dataset	Precision	Recall	F1	F1 [in paper]
CoNLL 03	90.03	91.46	90.79	92.22

Sarvesh Parab

Thank You

Any Questions?