Exploring Cross-sentence Contexts for Named Entity Recognition with BERT

Jouni Luoma, Sampo Pyysalo
TurkuNLP group, University of Turku, Turku, Finland

Year of Publication :- 2020
Number of Citations :- 36

Contents

Introduction

- Named entity recognition can be addressed as predicting the category of an entity.
- Named entity recognition (NER) approaches have evolved through various methodological phases.
  - Rule / Knowledge based approaches
  - Manually engineered features
    - These features contained information from document, dataset and external sources.
    - Cross sentence information[1]
    - Dense representation of text such as word, character, string and subword embeddings [2]
  - Deep learning approaches

---

[1] Krishnan, V. , Manning, C. D. An effective two-stage model for exploiting non-local dependencies in named entity recognition. in Proceedings of the 21st international conference on computational linguistics and 44th annual meeting of the association for computational linguistics (2006), 1121–1128.
[2] Collobert, R. et al. Natural language processing (almost) from scratch. Journal of machine learning research 12, 2493–2537 (2011).

Related Works

Existing Deep learning approaches

- State of the art Deep learning Approaches use representations for one sentence at a time.
- Deep learning approaches with contexual reresentations.
  - Maximal Document context provided by data [3].
  - Character level word embedding using biLSTM [4]
  - Pretraining bi directional transformer for single sentence task [5]
- These deep learning approaches for consider one sentence at a time even though models like BERT[3] can support input multiple sentences at a time with the supported window size.

---

[3] Devlin, J., Chang, M.-W., Lee, K. , Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018).
[4] Akbik, A., Blythe, D. , Vollgraf, R. Contextual string embeddings for sequence labeling. in Proceedings of the 27th international conference on computational linguistics (2018), 1638–1649.
[5] Baevski, A., Edunov, S., Liu, Y., Zettlemoyer, L. , Auli, M. Cloze-driven pretraining of self-attention networks. arXiv preprint arXiv:1903.07785 (2019).

Related Works

Approaches with broader context

- There is a chance that the information to determine the category might be found elsewhere.
- Following approaches try broader context.
  - Keep in memory and pooling of contexual representation for a word and use as a word feature [6] [7]
  - Sliding window approach to have some part of the previous input window [8]
- The above approaches use broader context than a sentence but, the cross sentence context is not analyzed.

---

[6] Akbik, A., Bergmann, T. , Vollgraf, R. Pooled contextualized embeddings for named entity recognition. in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (2019), 724–728.
[7] Luo, Y., Xiao, F. , Zhao, H. Hierarchical contextualized representation for named entity recognition. in Proceedings of the AAAI conference on artificial intelligence 34 (2020), 8441–8448.
[8] Wu, S. , Dredze, M. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. arXiv preprint arXiv:1904.09077 (2019).

Models and Data

- 5 different models for the 5 languages tested.
  - BERTje base, Cased for Dutch
  - BERT-Large, Cased (Whole Word Masking) for English
  - FinBERT base, Cased for Finnish
  - German BERT, Cased for German
  - BETO, Cased for Spanish
- For Spanish, multilingual BERT is also used for comparison.
- CoNLL02 and CoNLL03 datasets for other languages except Finnish (PER, ORG, LOC and MISC tagged).
- For Finnish two recently published named entity recognition corpora [9] [10]are used (PER, LOC, ORG, PROD, EVENT and DATE).

---

[9] Ruokolainen, T., Kauppinen, P., Silfverberg, M. , Lindén, K. A Finnish news corpus for named entity recognition. Language Resources and Evaluation 54, 247–272 (2020).
[10] Luoma, J., Oinonen, M., Pyykönen, M., Laippala, V. , Pyysalo, S. A broad-coverage corpus for Finnish named entity recognition. in Proceedings of the Twelfth Language Resources and Evaluation Conference (2020), 4615–4624.

Method

- Pre-trained BERT model with its last layer followed by a single dense layer is used.
- Three input sentence representations were experimented.
    - One sentence per input window.
    - Including following sentences.
    - Including preceding and following sentences.

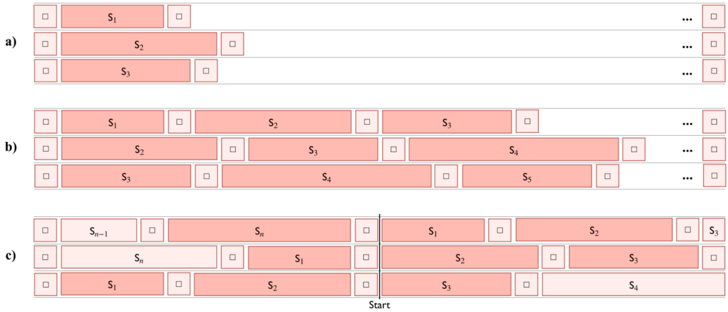

Figure: Illustration of various input representations for sequence labelling tasks. a) One sentence per example (Single), b) including following sentences (First, CMV), c) including preceding and following sentences (Sentence in context) [1]

1

Method

Context and CMV

- Constructing inputs in accordance with (b) and (c) allow the same sentences from the original data occur in different positions and with varying (sizes of) left and right contexts in different samples
- There are two variations in combining the results of multiple predictions at different contexts.
  - The first approach is to assign labels and then take a majority vote of the assigned labels.
  - The second approach is to add together the softmax probabilities of predictions in different contexts, and then take the argmax of the sum.

- Maximum length of the input sequence 512 is used.
- Hyparameters were tuned using exhaustive grid search within the parameters below.
  - Learning Rate: 2e-5, 3e-5 and 5e-5
  - Batch Size: 2,4,8,16
  - Epochs : 1,2,3,4

Results

Comparing languages and methods

	Precision	Recall	F1	F1 train+dev
English, CMV	93.06 (0.25)	**93.78** (0.08)	93.42 (0.12)	93.57 (0.33)
English, First	**93.15** (0.15)	93.73 (0.04)	**93.44** (0.06)	**93.74** (0.25)
English, Single	91.12 (0.25)	92.28 (0.23)	91.70 (0.24)	91.94 (0.15)
Dutch, CMV	**93.12** (0.26)	93.26 (0.18)	93.19 (0.21)	**93.49** (0.23)
Dutch, First	93.03 (0.65)	**93.38** (0.38)	**93.21** (0.51)	93.39 (0.26)
Dutch, Single	91.57 (0.35)	91.49 (0.41)	91.53 (0.37)	91.92 (0.30)
Finnish, CMV	**92.91** (0.18)	**94.42** (0.13)	**93.66** (0.13)	**93.78** (0.26)
Finnish, First	92.56 (0.14)	94.24 (0.08)	93.39 (0.10)	93.65 (0.26)
Finnish, Single	90.74 (0.10)	92.11 (0.24)	91.42 (0.16)	91.97 (0.21)
German, CMV	**86.91** (0.31)	**84.38** (0.32)	**85.63** (0.30)	**87.31** (0.27)
German, First	86.37 (0.39)	84.07 (0.10)	85.21 (0.22)	86.91 (0.11)
German, Single	85.55 (0.20)	81.81 (0.31)	83.64 (0.21)	85.67 (0.25)
Spanish, CMV	**87.80** (0.25)	**87.98** (0.18)	**87.89** (0.21)	**87.97** (0.21)
Spanish, First	86.71 (0.31)	87.41 (0.28)	87.06 (0.28)	87.27 (0.25)
Spanish, Single	87.43 (0.53)	87.90 (0.34)	87.66 (0.43)	87.52 (0.41)
S-mBERT, CMV	**87.25** (0.50)	**88.67** (0.46)	**87.95** (0.47)	**88.32** (0.26)
S-mBERT, First	86.92 (0.40)	87.88 (0.44)	87.40 (0.42)	87.54 (0.25)
S-mBERT, Single	87.19 (0.28)	87.81 (0.26)	87.50 (0.26)	87.57 (0.29)

Figure: NER results for different methods and languages (standard deviation in parentheses) [2]

Model	Our F1	Our F1 (t+d)	Current BERT	Current SOTA
English	93.44	**93.74**	93.47 (Liu et al., 2019b)	93.5 (Baevski et al., 2019)
Dutch	93.21	**93.49**	90.94 (Wu and Dredze, 2019)	92.69 (Straková et al., 2019)
Finnish	93.66	**93.78**	93.11 (Luoma et al., 2020)	93.11 (Luoma et al., 2020)
German	85.63	87.31	82.82 (Wu and Dredze, 2019)	**88.32** (Akbik et al., 2018)
Spanish	87.89	87.97	88.43 (Cañete et al., 2020)	**89.72** (Conneau et al., 2020)
Spanish, mBERT	87.95	88.32	88.43 (Cañete et al., 2020)	**89.72** (Conneau et al., 2020)

Table 3: NER result comparison to the state of the art.

- When considering a sentence, it was observed the NER performance was high when the sentence was near middle than being near to one end of the input window.
- Experimental Results show that NER performance benefits significantly using cross sentence context.
- Able this establishes a new state-of-the-art result for three languages, English, Dutch and Finnish

Conclusion

- A method to explore and make use of cross sentence context is presented for NER task.
- NER performance has a significant improvement for three languages when using the cross sentence context.

References I

1. Krishnan, V. & Manning, C. D. *An effective two-stage model for exploiting non-local dependencies in named entity recognition.* in *Proceedings of the 21st international conference on computational linguistics and 44th annual meeting of the association for computational linguistics* (2006), 1121–1128.
2. Collobert, R. *et al.* Natural language processing (almost) from scratch. *Journal of machine learning research* **12,** 2493–2537 (2011).
3. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
4. Akbik, A., Blythe, D. & Vollgraf, R. *Contextual string embeddings for sequence labeling.* in *Proceedings of the 27th international conference on computational linguistics* (2018), 1638–1649.

References II

5. Baevski, A., Edunov, S., Liu, Y., Zettlemoyer, L. & Auli, M. Cloze-driven pretraining of self-attention networks. *arXiv preprint arXiv:1903.07785* (2019).
6. Akbik, A., Bergmann, T. & Vollgraf, R. *Pooled contextualized embeddings for named entity recognition.* in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (2019), 724–728.
7. Luo, Y., Xiao, F. & Zhao, H. *Hierarchical contextualized representation for named entity recognition.* in *Proceedings of the AAAI conference on artificial intelligence* **34** (2020), 8441–8448.
8. Wu, S. & Dredze, M. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. *arXiv preprint arXiv:1904.09077* (2019).

References III

9. Ruokolainen, T., Kauppinen, P., Silfverberg, M. & Lindén, K. A Finnish news corpus for named entity recognition. *Language Resources and Evaluation* **54,** 247–272 (2020).

10. Luoma, J., Oinonen, M., Pyykönen, M., Laippala, V. & Pyysalo, S. *A broad-coverage corpus for Finnish named entity recognition.* in *Proceedings of the Twelfth Language Resources and Evaluation Conference* (2020), 4615–4624.

Thank You! Any Questions?