Leveraging Linguistic Structures for Named Entity Recognition with Bidirectional Recursive Neural Networks

Presenter: Peng-Hsuan Li

November 15, 2016

Motivation

Names

PERSON	People, including fictional
NORP	Nationalities or religious or political groups
FACILITY	Buildings, airports, highways, bridges, etc.
ORGANIZATION	Companies, agencies, institutions, etc.
GPE	Countries, cities, states
LOCATION	Non-GPE locations, mountain ranges, bodies of water
PRODUCT	Vehicles, weapons, foods, etc. (Not services)
EVENT	Named hurricanes, battles, wars, sports events, etc.
WORK OF ART	Titles of books, songs, etc.
LAW	Named documents made into laws

Numbers

DATE	Absolute or relative dates or periods
TIME	Times smaller than a day
PERCENT	Percentage (including %)
MONEY	Monetary values, including unit
QUANTITY	Measurements, as of weight or distance
ORDINAL	first, second
CARDINAL	Numerals that do not fall under another type

Named Entity Recognition

Input - word sequences:

the defense secretary Donald Rumsfeld

Output - located names and their types:

(1, 2, ORG)
(3, 5, PERSON)

Sequential Labeling NER

Sequence labels for words:

Begin, Inside, End, Outside, Single

Input - word sequences:

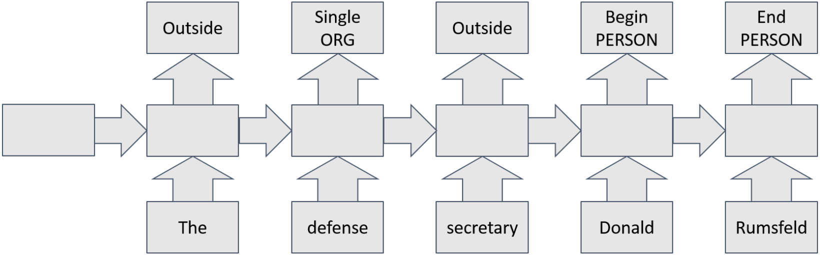
the defense secretary Donald Rumsfeld

Output - labeled word sequences:

the	defense	secretary	Donald	Rumsfeld
Outside	Single	Outside	Begin	End
	ORG		PERSON	PERSON

Note that this is interchangeable to the non-sequential representation.

Sequetial labeling NER with Recurrent Networks



Incorporate Linguistic Structures

Named entity chunks are in most cases actually linguistic phrases, especially noun phrases.

E.g. this prior knowledge should be useful:

"Donald Rumsfeld" is a meaningful phrase
"secretary Donald" is not a meaningful phrase

NER systems trained on only NER labels are hard to gain this knowledge implicitly.

Problem Statement

Leveraging Linguistic Structures for NER

Input:

- A segmented token (word) sequence
- The parse tree of the sequence

External Resources:

- GloVe embeddings trained on 840 billion tokens on the web

Output:

- Located names and their types of the sequence

Related Work

Latest Work on NER

2003	CoNLL 2003 shared task [Tjong Kim Sang and De Meulder, 2003]
2009	Joint parsing and NER with CRF on OntoNotes [Finkel and Manning, 2009]
2011	Comparable pure NN on CoNLL 2003 [Collobert et al., 2011]
2013	OntoNotes 5.0 [Pradhan et al., 2013]
2014	Joint NER, linking, coreference with CRF on OntoNotes 5.0 [Durrett and Klein, 2014]
2015	Joint NER and linking with CRF on CoNLL 2003 [Luo et al., 2015]
	Character-level embedding for Portuguese and Spanish NER with CNN [dos Santos and Guimaraes, 2015]
2016	The state-of-the-art system with LSTM-CNN on CoNLL 2003 and OntoNotes 5.0 [Chiu and Nichols, 2016]

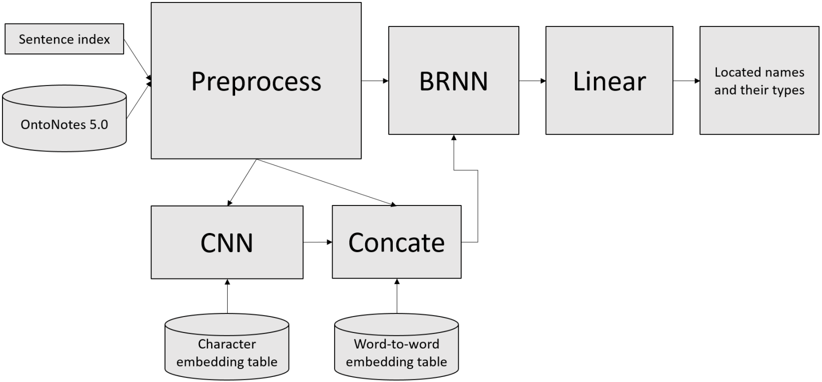
Other Important Related Work

2010	Parsing with recursive networks [Socher et al., 2010]
2013	Opinion expression extraction with recursive networks [Irsoy and Cardie, 2013]
	Sentiment analysis with recursive netoworks [Socher et al., 2013]
2014	GloVe word embedding [Pennington et al., 2014]
2015	Sentiment analysis with recursive LSTM netoworks [Tai et al., 2015]
2016	Character-aware neural language models [Kim et al., 2016]

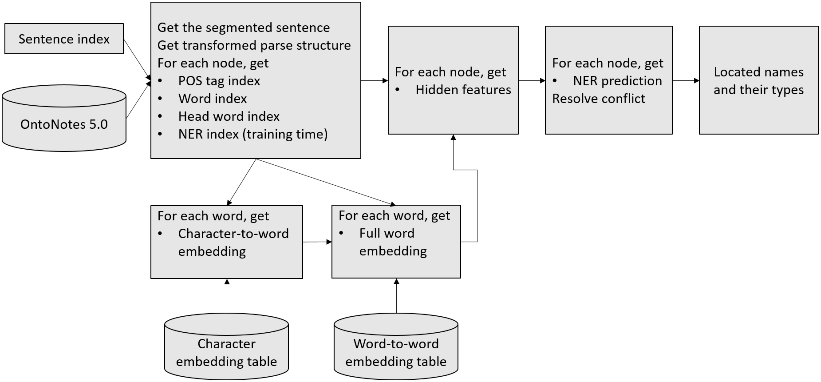
Method

System Overview

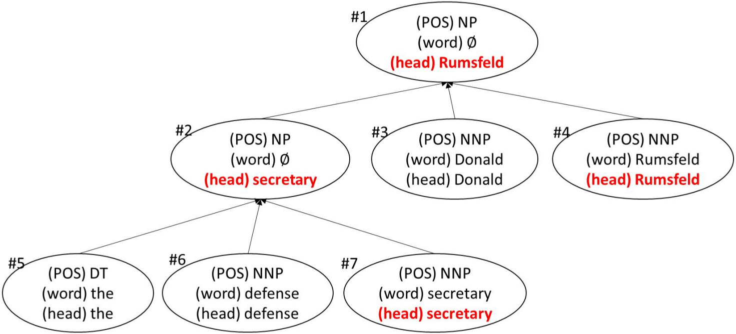
System Overview



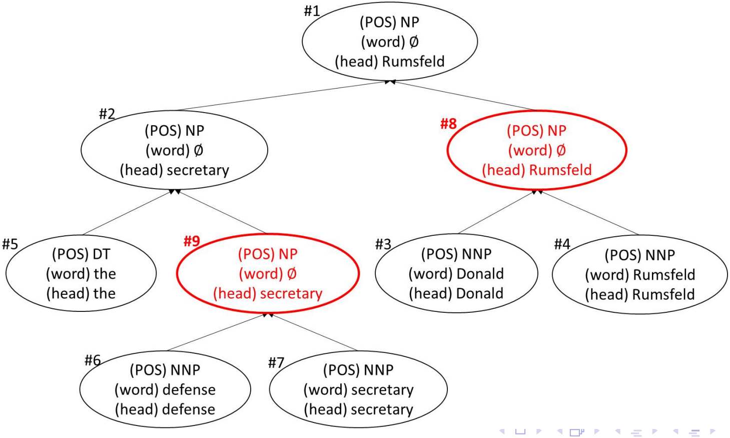
System Overview



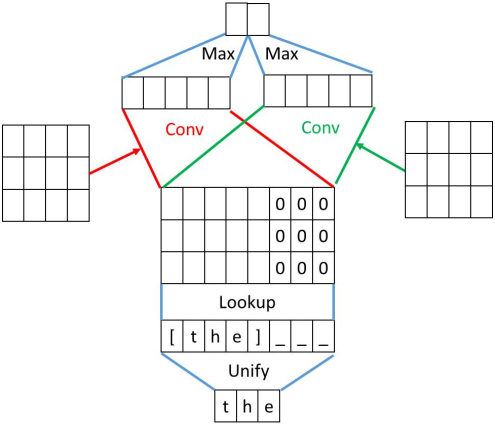
Preprocess: Parses



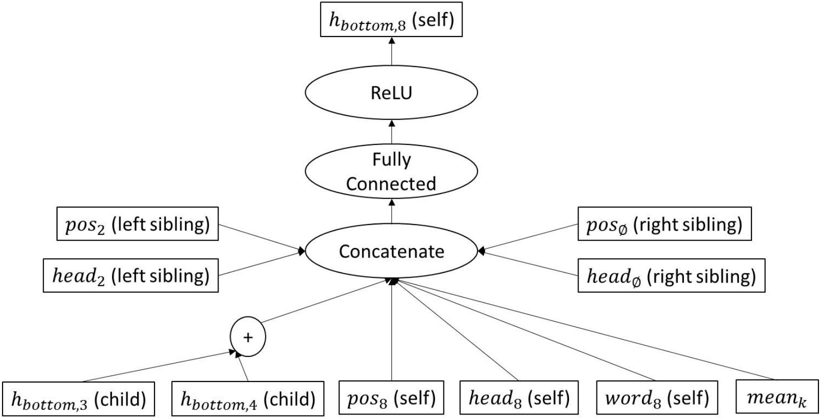
Preprocess: Parses



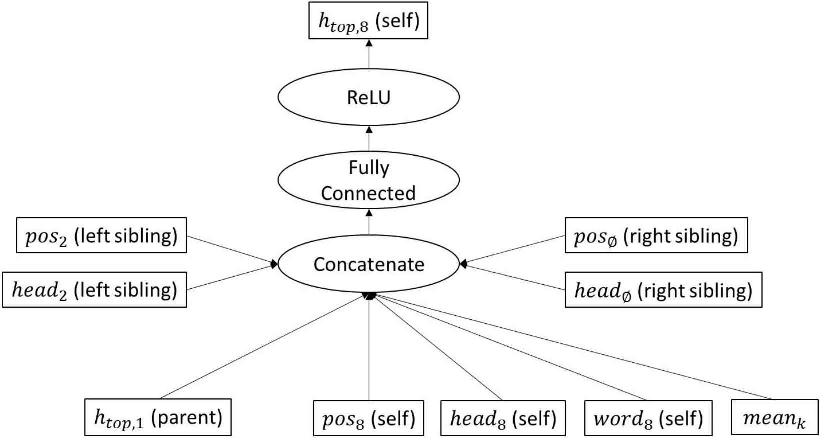
CNN: Character-to-word Embedding



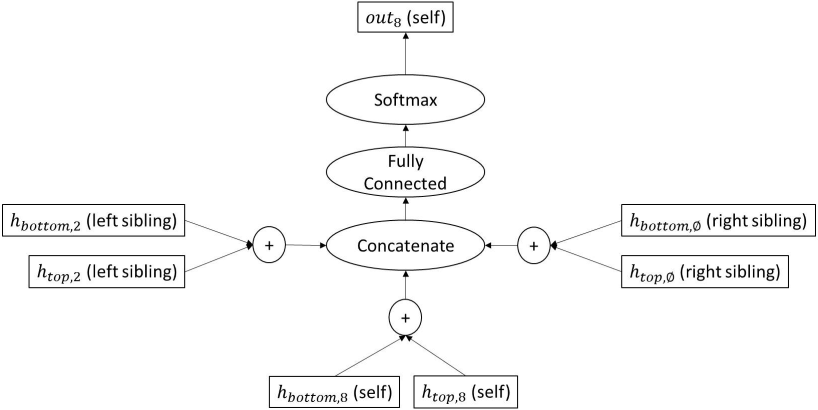
BRNN: Bottom-up Hidden Layer Applied to Node #8



BRNN: Top-down Hidden Layer Applied to Node #8



Linear: The Output Layer Applied to Node #8



Evaluation

OntoNotes 5.0

18 types of named entities

$\sim$70,000 sentences

$\sim$100,000 name entities

Data sources:

- BC (broadcast conversation)
- BN (broadcast news)
- MZ (magazine)
- NW (newswire)
- TC (telephone conversation)
- WB (blogs and newsgroups)

Trials

CoNLL-2012 train/validation/test split

10 successful trials per main model
2 successful trials per other model

Results - Whole Dataset

	Validation			Test		
	Precision	Recall	F1	Precision	Recall	F1
BRNN	84.63	85.47	85.05 (0.13)	86.17	86.92	86.54 (0.36)
BRNN-CNN	84.62	**85.77**	**85.19 (0.17)**	**86.24**	**87.09**	**86.67 (0.13)**
BRNN-gold [1]	**85.61**	**87.88**	**86.72 (0.16)**	**87.80**	**89.31**	**88.54 (0.17)**
RNN	84.05	84.70	84.40 (0.04)	85.75	86.10	85.91 (0.13)
BRNN-updown	**84.75**	85.45	85.12 (0.28)	86.20	86.80	86.49 (0.15)
BRNN-concat	84.35	85.45	84.89 (0.13)	86.10	86.90	86.50 (0.41)
rNN	83.10	83.70	83.38 (0.58)	84.45	84.40	84.40 (0.41)
[Durrett and Klein, 2014]	-	-	-	85.22	82.89	84.04
[Chiu and Nichols, 2016][2]	-	-	-	85.99	86.36	86.17 (0.22)
[Chiu and Nichols, 2016]	-	-	-	-	-	86.41 (0.22)

Results - Different Sources

Model	BC	BN	MZ	NW	TC	WB
Test set size (# tokens)	32576	23557	18260	51667	11015	19348
Test set size (# entities)	1697	2184	1163	4696	380	1137
[Finkel and Manning, 2009]	78.66	87.29	82.45	85.50	67.27	72.56
[Durrett and Klein, 2014]	78.88	87.39	82.46	87.60	72.68	76.17
[Chiu and Nichols, 2016]	85.23	89.93	**84.45**	88.39	72.39	78.38
BRNN	85.17	**90.37**	83.84	**88.85**	74.34	**81.32**
BRNN-CNN	**85.45**	90.19	84.39	88.48	**75.03**	80.93

Results - Significance

	Samples	Sample F1 Mean	Sample F1 Deviation
BRNN-CNN	10	86.67	0.13
[Chiu and Nichols, 2016]	10	86.41	0.22

By one-tailed Welch's t-test, the null hypothesis that BRNN-CNN doesn't have a higher population mean is rejected with p-value 0.003654 (99% confidence level).

Time Plan

Time Plan

2016-12	02	EACL long papers: rejection notification
	16	EACL short papers: submission due
		Develop/incorporate a parser into BRNN
		Evaluate on OntoNotes 5.0 Chinese
2017-01		Evaluate on CoNLL 2003
		Analyze output

Reference I

Chiu, J. P. and Nichols, E. (2016).
Named entity recognition with bidirectional lstm-cnns.
Transactions of the Association for Computational Linguistics, 4:357–370.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011).
Natural language processing (almost) from scratch.
Journal of Machine Learning Research, 12(Aug):2493–2537.

dos Santos, C. and Guimaraes, V. (2015).
Boosting named entity recognition with neural character embeddings.
In *Proceedings of the Fifth Named Entity Workshop, joint with 53rd ACL and the 7th IJCNLP*, pages 25–33.

Durrett, G. and Klein, D. (2014).
A joint model for entity analysis: Coreference, typing, and linking.
Transactions of the Association for Computational Linguistics, 2:477–490.

Finkel, J. R. and Manning, C. D. (2009).
Joint parsing and named entity recognition.
In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 326–334. Association for Computational Linguistics.

Reference II

Irsoy, O. and Cardie, C. (2013).
Bidirectional recursive neural networks for token-level labeling with structure.
In *NIPS Deep Learning Workshop*.

Kim, Y., Jernite, Y., Sontag, D., and Rush, A. M. (2016).
Character-aware neural language models.
In *Thirtieth AAAI Conference on Artificial Intelligence*.

Luo, G., Huang, X., Lin, C.-Y., and Nie, Z. (2015).
Joint named entity recognition and disambiguation.
In *Proc. EMNLP*, pages 879–888.

Pennington, J., Socher, R., and Manning, C. D. (2014).
Glove: Global vectors for word representation.
In *EMNLP*, volume 14, pages 1532–1543.

Pradhan, S., Moschitti, A., Xue, N., Ng, H. T., Björkelund, A., Uryupina, O., Zhang, Y., and Zhong, Z. (2013).
Towards robust linguistic analysis using ontonotes.
In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 143–152.

Reference III

Socher, R., Manning, C. D., and Ng, A. Y. (2010).
Learning continuous phrase representations and syntactic parsing with recursive neural networks.
In *Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop*.

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., and Potts, C. (2013).
Recursive deep models for semantic compositionality over a sentiment treebank.
In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642.

Tai, K. S., Socher, R., and Manning, C. D. (2015).
Improved semantic representations from tree-structured long short-term memory networks.
In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processingg*, pages 1556–1566.

Tjong Kim Sang, E. F. and De Meulder, F. (2003).
Introduction to the conll-2003 shared task: Language-independent named entity recognition.
In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 142–147. Association for Computational Linguistics.