Named Entity Recognition Based on Bidirectional Recursive Neural Networks

20 Minutes Presentation

Peng-Hsuan Lee

October 12, 2016

Introduction to Named Entity Recognition (NER)

Names and Numbers

Names:

Person, location, organization... etc.

Numbers:

Cardinal, ordinal, data... etc.

Extracting Names

Unstructured texts:

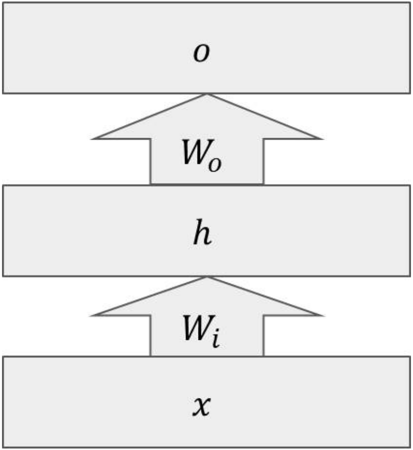
... the defense secretary Donald Rumsfeld ...

Sequential labeling:

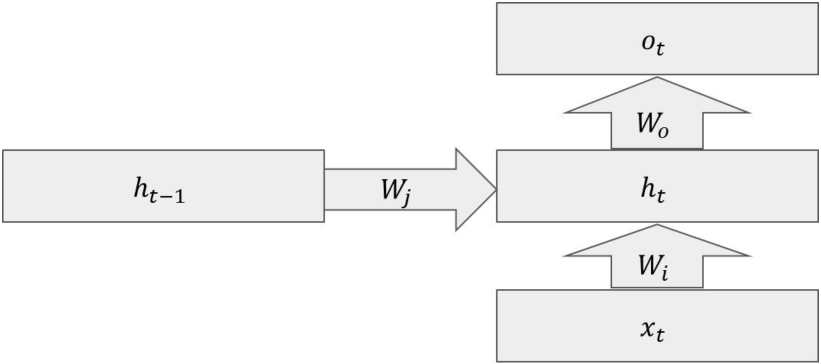
... the(X) defense(ORG_start) secretary(X)
Donald(PERSON_start) Rumsfeld(PERSON_continue) ...

Introduction to Recursive Neural Network (RNN)

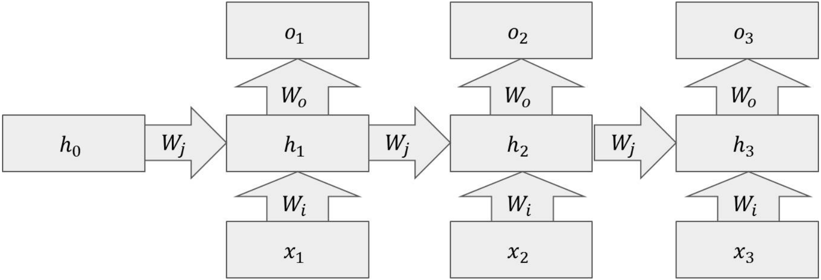
A Vanilla Network



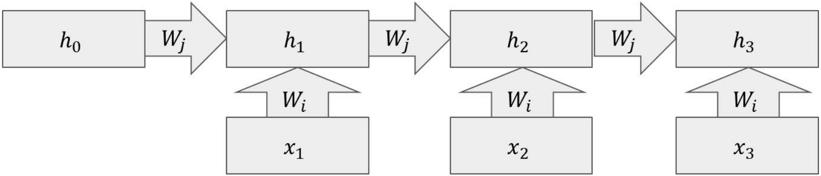
A Recurrent Network



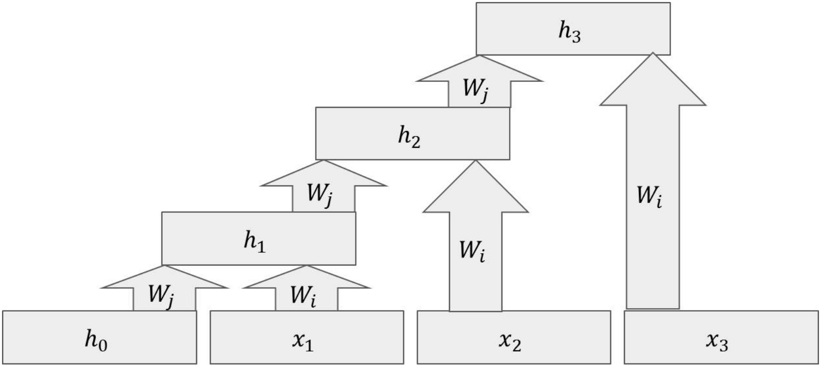
A Recurrent Network - Expanded View



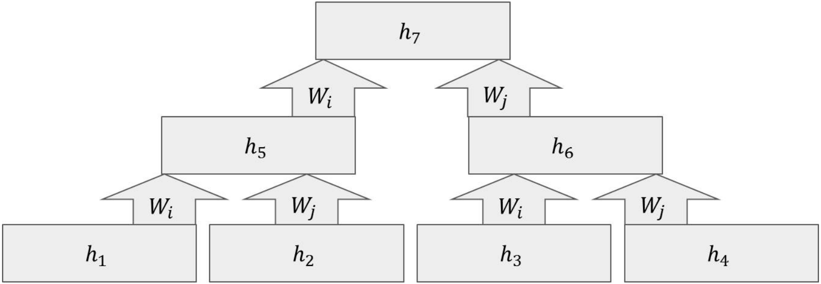
A Recurrent Network - Simplified View



A Recurrent Network - Tree-ed View

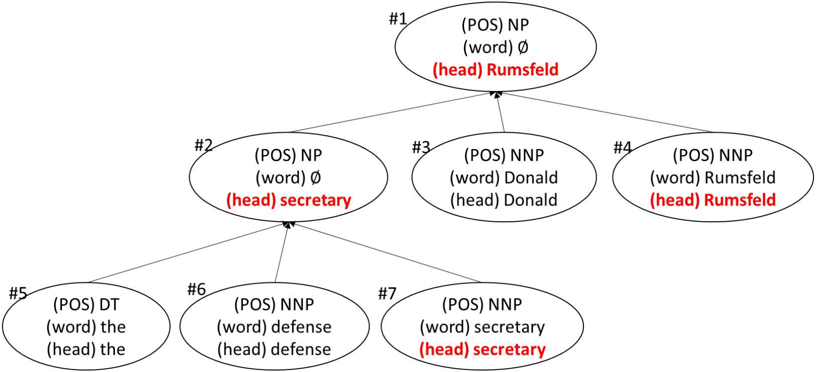


A Recursive Network

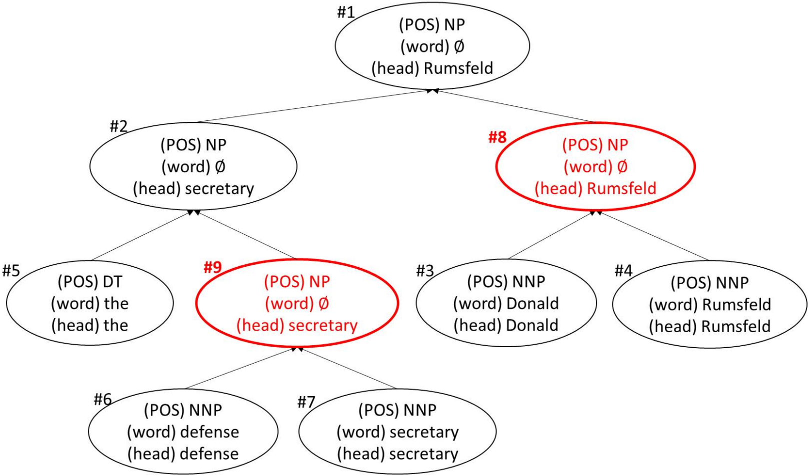


Model

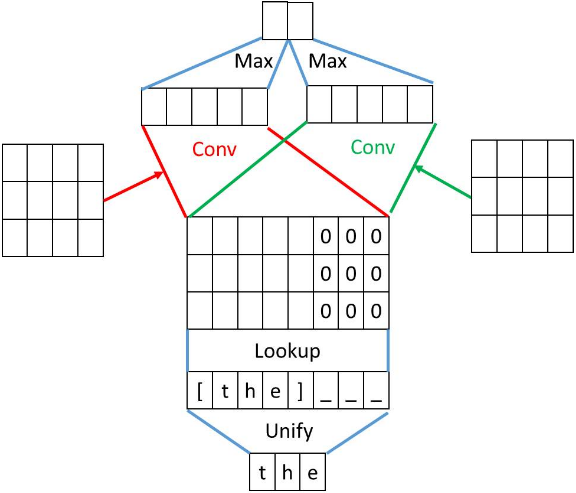
A Constituency Parse



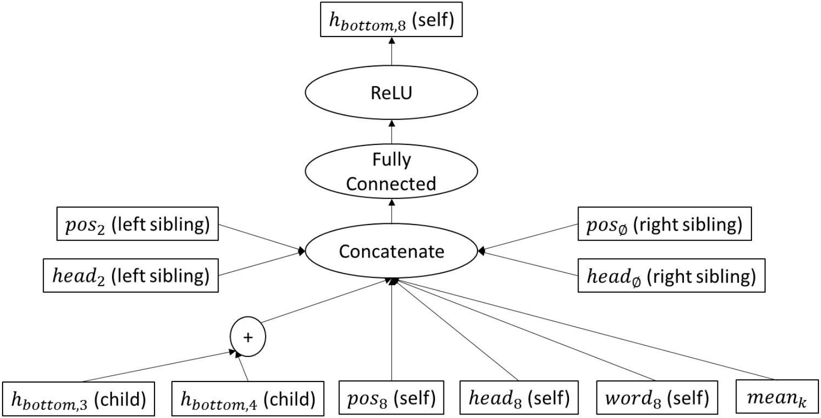
Transformed Parse



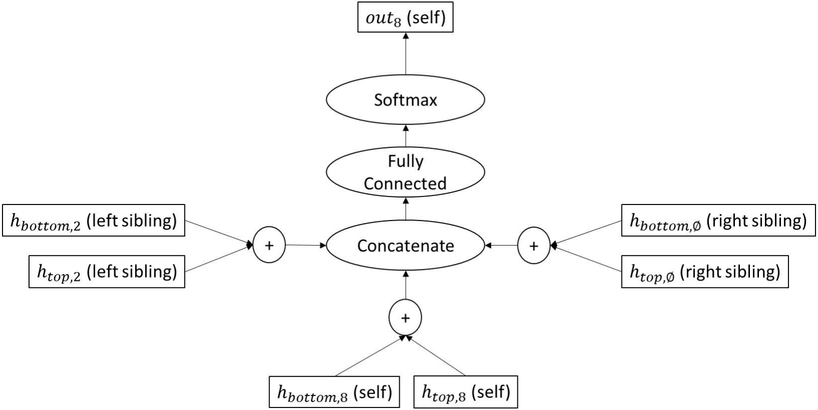
An Example Character-level CNN



Bottom-up Hidden Layer Applied to Node #8



The Output Layer Applied to Node #8



Experiment

OntoNotes 5.0

18 types of named entities

∼70,000 sentences

∼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

Related Work

NER

[Collobert et al., 2011] NLP from scratch

[Passos et al., 2014] 1^{st} OntoNotes NER

[Durrett and Klein, 2014] CRF + joint training for NER

[dos Santos and Guimaraes, 2015] CNN for Portuguese/Spanish NER

[Chiu and Nichols, 2016] LSTM-CNN for NER

RNN

[Socher et al., 2010] RNN for parsing

[Socher et al., 2013] RNN for sentiment analysis

[Tai et al., 2015] Tree-LSTM for sentiment analysis

[Irsoy and Cardie, 2013] Token-level bidirectional RNN for Opinion expression identification

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*.

Passos, A., Kumar, V., and McCallum, A. (2014).
Lexicon infused phrase embeddings for named entity resolution.
In *Proceedings of the Eighteenth Conference on Computational Language Learning*, pages 78–86.

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.