A Named Entity Recognition Shootout for German

Martin Riedl and Sebastian Padó

Institut für maschinelle Sprachverarbeitung (IMS), Universität Stuttgart, Germany {martin.riedl, pado}@ims.uni-stuttgart.de

JIntroductionsk: Named Entity Recognition (NER)Recognition of proper names, e.g. locations, persons, organizations etc.TheBritishGovernmentsentCaptainArthurPhilliptoestablishacolonyinNewSouthWalesDB-ORGI-ORGOB-PERI-PEROOOOOB-LOCI-LOC							3 Exp. 1: Which method performs best on the contemporary datasets?					
 B-ORG I-ORG B-PER I-PER I-PER O O O O O B-LOC I-LOC I-LOC I-LOC Research Questions How can we build state-of-the-art performing German NER systems trained on big data (contemporary data)? small data (historic data)? What are the performance differences between: <u>Traditional</u> CRF: established, fast, feature engineering, work with few amounts of training data BiLSTM+CRF: representation learning, no feature engineering needed, long 						Model CoNLL GermEval P R F1 P R F1 StanfordNER 74.18 72.50 73.33 80.13 65.43 72.04 GermaNER 85.99 73.78 79.37 82.72 71.19 76.52 BiLSTM-WikiEmb 87.67 78.79 82.99* 83.07 80.62 81.83* BiLSTM-EuroEmb 79.92 72.14 75.83 76.48 73.54 74.98 BiLSTM outperforms CRFs due to higher recall Biles recall Biles recall Biles recall Biles recall Biles recall						
distance dep otain best pra	actice for buil	lding NER sy	ge amounts of training stems			4	per	p.2 forn lacr	nan	ce v	vit	hin
Datasets	Contempor	arv Texts			Model	Train CoNLL	72.12	Test dat GermEval 48.82	LFT 39.72	ONB 46.36 52.21		
	CoNLL 2003 [1]	GermEval 2014 [2] Wikipedia	Historic Friedrich Teßmann Library (LFT) [3] Newspaper	Austrian National Library (ONB) [3] Newspaper			a Stanfor R NER	GermEval LFT ONB CoNLL	35.25 34.09 79.37	72.09 35.00 33.96 60.40 76.37	45.22 67.26 42.95 46.53 48.05	52.21 52.77 72.42
Text			- · · · · · · · · · · · · · · · · · · ·					GermEval				53.93 54.95
Time	~2000	2001-2004 450,000	1926 87,000	1710-1873 35,000			Emb Germa	GermEval LFT ONB CoNLL GermEval		45.82 47.19 66.51 82.93	69.18 48.41 49.28 55.99	
Time Tokens Methods • CRF-ba • S	~2000 220,000 sed methods tanfordNER	450,000 s: [4]: CRF + sta		1710-1873 35,000				LFT ONB CoNLL	46.56 82.99 78.15 57.27 51.42 75.83	47.19 66.51	48.41 49.28 55.99 68.47 49.35 45.30 52.15 69.62	54.95 56.38 73.31 58.79

[1] E. F. Tjong Kim Sang and F. De Meulder. 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In Proc. of CoNLL-2003. Edmonton, Canada, pages 142-147 [2] D. Benikova, C. Biemann, and M. Reznicek. 2014. NoSta-D Named Entity Annotation for German: Guidelines and Dataset. In Proceedings of LREC. Reykjavik, Iceland, pages 2524-2531 [3] C. Neudecker. 2016. An open corpus for named entity recognition in historic newspapers. In Proc. of LREC. Portoro, Slovenia, pages 4348-4352 [4] J. R.Finkel, T. Grenager, and C. Manning. 2005. Incorporating non-local information into information extraction systems by Gibbs sampling. In Proc. of ACL. Ann Arbor, MI, USA, pages 363-370 [5] D. Benikova, S. Muhie Yimam, and C. Biemann. 2015. GermaNER: Free Open German Named Entity Recognition Tool. In Proc. of GSCL. Essen, Germany, pages 31-38 [6] X. Ma and E. Hovy. 2016. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs- CRF. In Proc. of ACL. Berlin, Germany, pages 1064-1074

Same setup as Exp. 2 but with transfer learning and considering only the BiLSTM-based method

			BiLSTM-WikiEmb				BiLSTM-EuroEmb				
Train	Transfer	CoNLL	GermEval	LFT	ONB	CoNLL	GermEval	LFT	ONB		
CoNLL	GermEval	78.55	82.93	55.28	64.93	72.23	75.78	51.98	61.74		
CoNLL	LFT	62.80	58.89	72.90	67.96	56.30	51.25	70.04	65.65		
CoNLL	ONB	62.05	57.19	59.43	76.17	55.82	49.14	54.19	73.68		
GermEval	CoNLL	84.73 [†]	72.11	54.21	65.95	78.41	63.42	52.02	59.28		
GermEval	LFT	67.77	69.09	74.33 [†]	70.57	55.83	57.71	72.03	70.36		
GermEval	ONB	72.15	73.18	62.52	76.06	64.05	64.20	57.12	78.56		

Contemporary corpora: minor improvements GermEval: 82.93 → 82.93 CoNLL: 82.99 → 84.73 Historic corpora: major improvements LFT: 69.62 → 73.44 ONB: $70.46 \rightarrow 78.56$

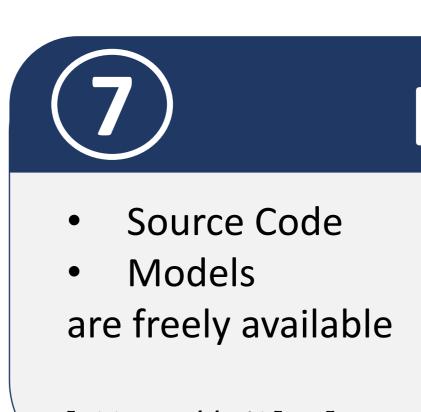
LSTM + CRF based models outperform traditional CRF if:

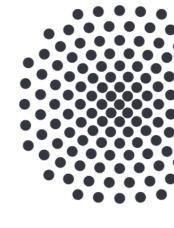
- lots of training data is available
- if transfer learning is used
- Usage of character- and substring-based embeddings
- (FastText) solves OOV issues

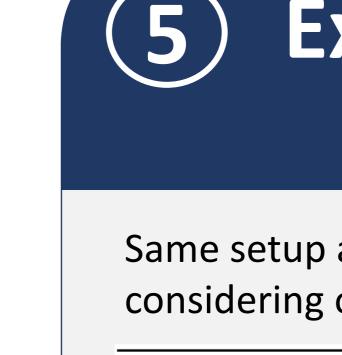
Future

(6)

- Learn multilingual models
- Analyze features learned by LSTM









Universität Stuttgart

Exp. 3: Can Transfer Learning help?

Conclusion & Future

Download



https://github.com/riedlma/sequence_tagging

56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia