ℓ_1 Regularization of Word Embeddings for Multi-Word Expression Identification

Gábor Berend

Institute of Informatics, University of Szeged

berendg@inf.u-szeged.hu

Overview

- Multi-word expressions (MWEs) are lexical items with spaces (called *catenas* in linguistics)
- We conducted controlled experiments to compare the effects of various **word representations** and **classification algorithms** for MWE identification

MWE dataset used

- WIKI50 corpus Vincze et al. (2011) (cca. 114K tokens and 4.4K sentences)
- Multiple types of MWEs

MWE type	Example
Noun compounds	*black box*
Adjectival compounds	*monkey styled*
Verb-Particle Constructions (VPC)	*went on*
Light-Verb Constructions (LVC)	*opens fire*
Idioms	*caught the eye of*
Other	*alter ego*
Location NE	*Sierra Leone*
Person NE	*Sir Elton John*
Organization NE	*Major Indoor Soccer League*

Compared models

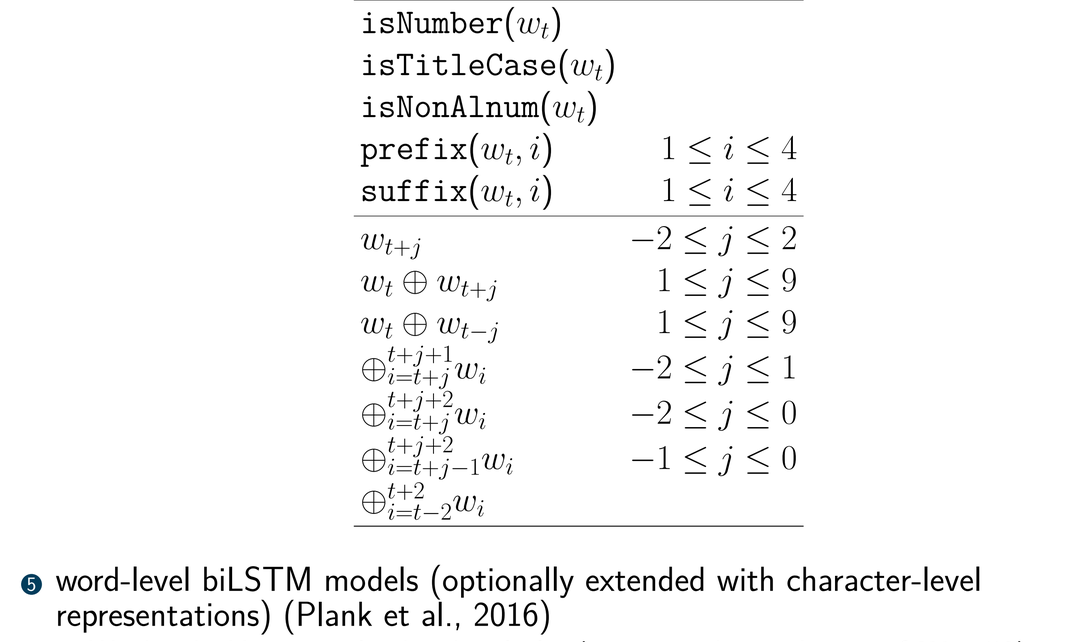
1. Linear CRF model based on pre-trained dense word embeddings
   - Glove, polyglot, skip-gram, CBOW embeddings trained on English Wikipedia
2. Linear CRF model based on ℓ_1-regularized word embeddings
   - For an embedding matrix $X \in \mathbb{R}^{d \times |V|}$ we solve for $\min_{D \in \mathcal{C}, \alpha} \|X - D\alpha\|_F^2 + \lambda \|\alpha\|_1$
     - $\mathcal{C}$ is the convex set of $\mathbb{R}^{d \times k}$ matrices with column norms ≤ 1, and α contains the sparse coefficients for word forms
   - For word i features are derived by $\phi_{sparse}(i) = \{j : 1 | \alpha_x[j,i] > 0 \wedge 1 \leq j \leq k\}$
3. Linear CRF model based on Brown-cluster prefixes of word forms as features
4. CRF model based on "traditional" feature set ($\oplus$ denoting concatenation)

Feature template	
$\texttt{isNumber}(w_t)$	
$\texttt{isTitleCase}(w_t)$	
$\texttt{isNonAlnum}(w_t)$	
$\texttt{prefix}(w_t, i)$	$1 \leq i \leq 4$
$\texttt{suffix}(w_t, i)$	$1 \leq i \leq 4$
w_{t+j}	$-2 \leq j \leq 2$
$w_t \oplus w_{t+j}$	$1 \leq j \leq 9$
$w_t \oplus w_{t-j}$	$1 \leq j \leq 9$
$\oplus_{i=t+j}^{t+j+1} w_i$	$-2 \leq j \leq 1$
$\oplus_{i=t+j}^{t+j+2} w_i$	$-2 \leq j \leq 0$
$\oplus_{i=t+j-1}^{t+j+2} w_i$	$-1 \leq j \leq 0$
$\oplus_{i=t-2}^{t+2} w_i$	

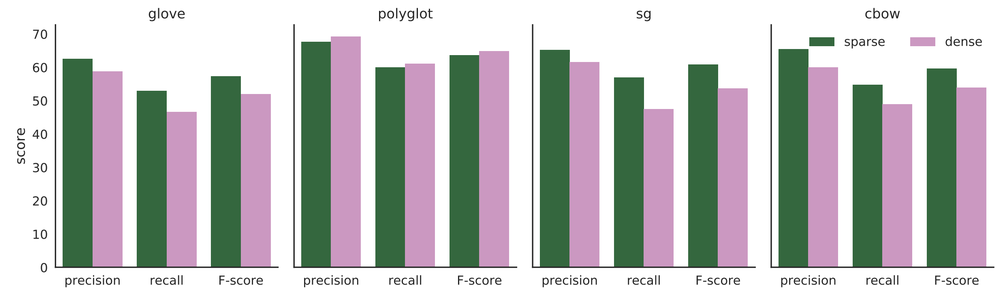
5. word-level biLSTM models (optionally extended with character-level representations) (Plank et al., 2016)
   - Word embeddings initialized with polyglot (important due to the size of WIKI50)

The effects of ℓ_1 regularization

- Compared the effects of relying on differently trained word embeddings (all trained on the same Wikipedia dumps)
- $d = 64, k = 1000$, sparsity ranging between 0.5% and 5%



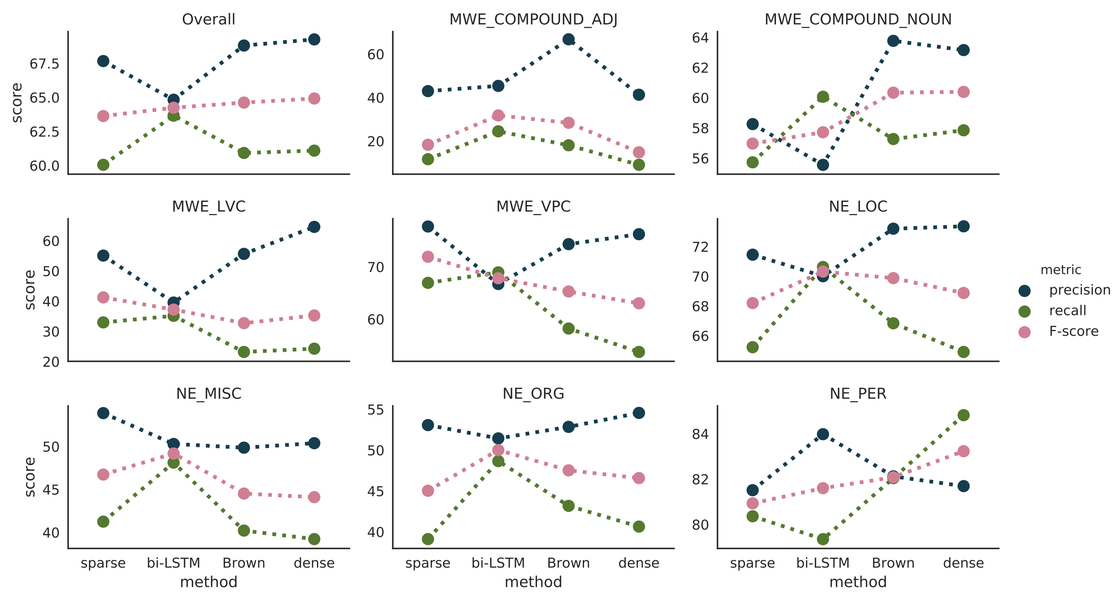
Dense versus sparse representation



Comparison of alternative models

method	Precision	Recall	F-score
FRw	62.91	18.04	28.04
FRwc	62.92	50.30	55.91
polyglot sparse ($\lambda = 0.3$)	67.65	60.03	63.61
bi-LSTM	64.81	63.64	64.22
Brown	68.79	60.90	64.60
polyglot dense	69.24	61.07	64.90

Detailed performance of alternative models



Extending biLSTM with character-level representation

- Improvements mostly on the NE categories

	biLSTM	biLSTM with char
Compound adj	**31.67**	6.32
Compound noun	57.70	**59.82**
LVC	**37.12**	33.74
VPC	**67.77**	61.17
Location	70.30	**72.98**
Misc	49.15	**51.66**
Organization	49.98	**53.55**
Person	81.59	**84.50**
Avg.	64.22	**66.48**

Conclusions

- biLSTM tends to produce more balanced results with respect to precision-recall (except for the `Persons`)
- Models employing sparse word representation perform best on verb-related MWE types (`VPC` and `LVC`)

References

G. Berend. Sparse coding of neural word embeddings for multilingual sequence labeling. *Transactions of the Association for Computational Linguistics*, 5:247–261, 2017. ISSN 2307-387X. URL https://transacl.org/ojs/index.php/tacl/article/view/1063.

B. Plank, A. Søgaard, and Y. Goldberg. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 412–418. Association for Computational Linguistics, 2016. URL http://anthology.aclweb.org/P16-2067.

V. Vincze, I. Nagy T., and G. Berend. Multiword expressions and named entities in the Wiki50 corpus. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, pages 289–295, Hissar, Bulgaria, September 2011. RANLP 2011 Organising Committee. URL http://aclweb.org/anthology/R11-1040.