Big Data Small Data, In Domain Out-of Domain, Known Word Unknown Word: The Impact of Word Representations on Sequence Labelling Tasks

Lizhen Qu, Gabriela Ferraro, Liyuan Zhou, Weiwei Hou,
Nathan Schneider & Timothy Baldwin

8 July 2015

Australian
National
University

Introduction

The Impact of Word Representations for Sequence Labellings Tasks

---

Distributed word representations, pre-trained from unlabelled data have become a mainstay of NLP, but. . . their quality depends on factors including:

- parameter optimisation
- the size of the training data
- the fit with the target application

Extrinsic evaluation on 4 labelling tasks under **FIX** experiment conditions:

- ✓ POS tagging
- ✓ Chunking
- ✓ NER
- ✓ MWE (multi-word expressions identification)

What we want to know about word embeddings

RQ1: Are word embeddings better than baseline approaches of one-hot unigram features and Brown clusters?

RQ2: To what degree can word embeddings reduce the amount of labelled data?

RQ3: What is the impact of updating word embeddings in sequence labelling tasks

RQ4: What is the impact of these word embeddings Out-of-Vocabulary items for *in-domain* and *out-of-domain* data?

RQ5: Are some word embeddings better than others in a sequence labelling context?

Experiments Setup

Fix the experiments conditions

- **Pre-trained word embeddings**
  - Corpora:
    - UMBC (Han et al.2013), 48.1GB
    - One Billion (Chelba et al.2013), 4.1GB
    - English Wikipedia, 49.6GB
  - Embedding dimensions:

    $d \in \{25, 50, 100, 200\}$
  - Context window size:

    $m \in \{1, 5, 10\}$

- **Selected Word Representations**
  - Brown clustering (Brown et al.1992)
  - CBOW (Mikolov et al.2013a),
  - Skipgram (Mikolov et al.2013b),
  - Global vectors (Pennington et al.2014)

Training, development and test data

	Training	Development	*In-domain* Test	*Out-of-domain* Test
POS tagging	WSJ Sec. 0-18	WSJ Sec. 19–21	WSJ Sec. 22–24	EWT
Chunking	WSJ	WSJ (1K sentences)	WSJ (CoNLL-00 test)	Brown
NER	Reuters (CoNLL-03 train)	Reuters (CoNLL-03 dev)	Reuters (CoNLL-03 test)	MUC7
MWE	EWT (500 docs)	EWT (100 docs)	EWT (123 docs)	—

RQ1: Word embeddings vs. one-hot unigram features vs. Brown clusters?

RQ2: To what degree can word embeddings reduce the amount of labelled data?

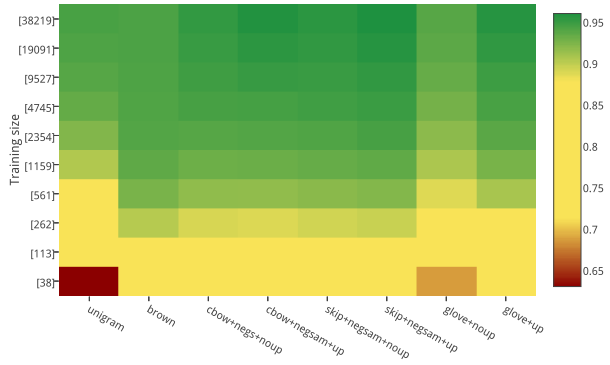


Figure : POS tagging (Acc)

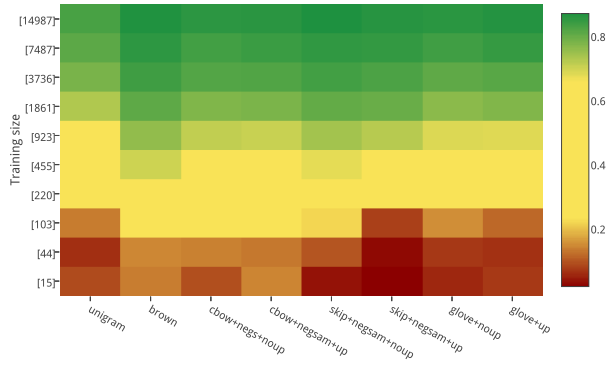


Figure : NER (F1)

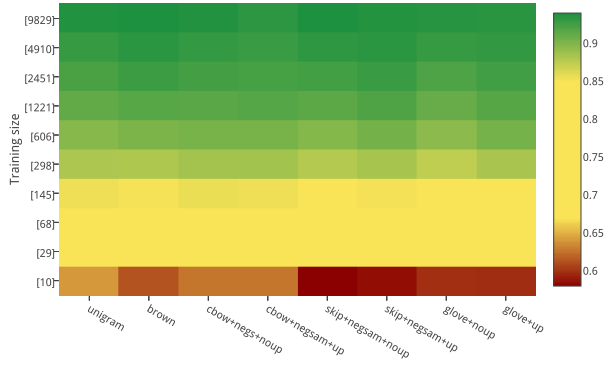


Figure : Chunking (F1)

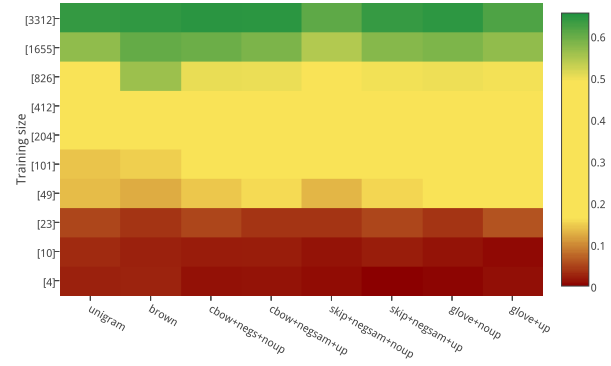


Figure : MWE (F1)

RQ3: : What is the impact of updating word embeddings geometrically over the vectors?

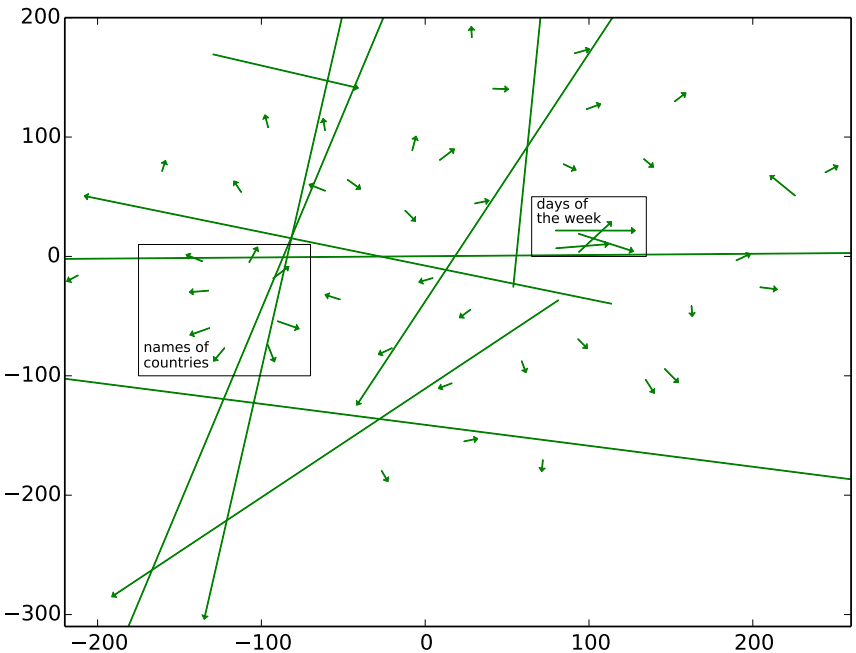


Figure : Chunking with SKIP-GRAM

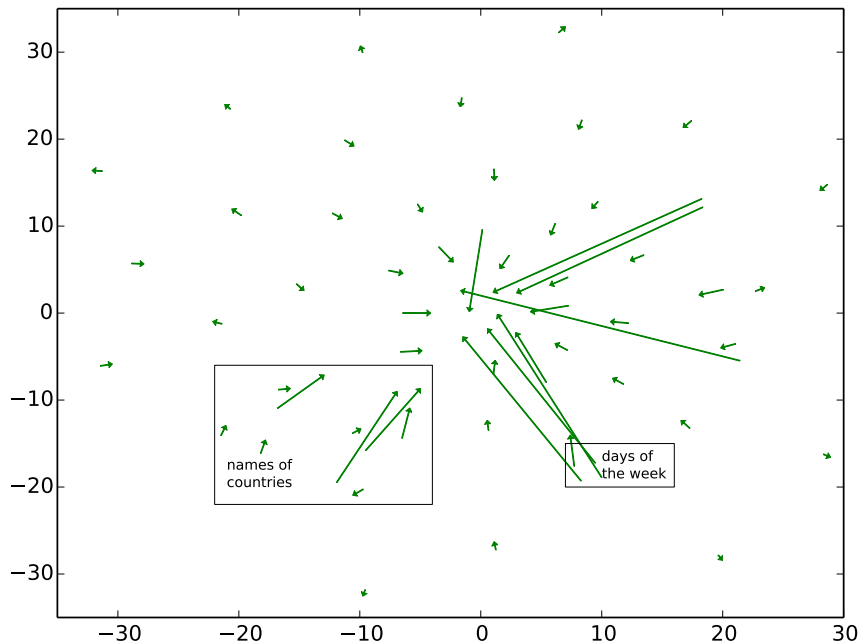
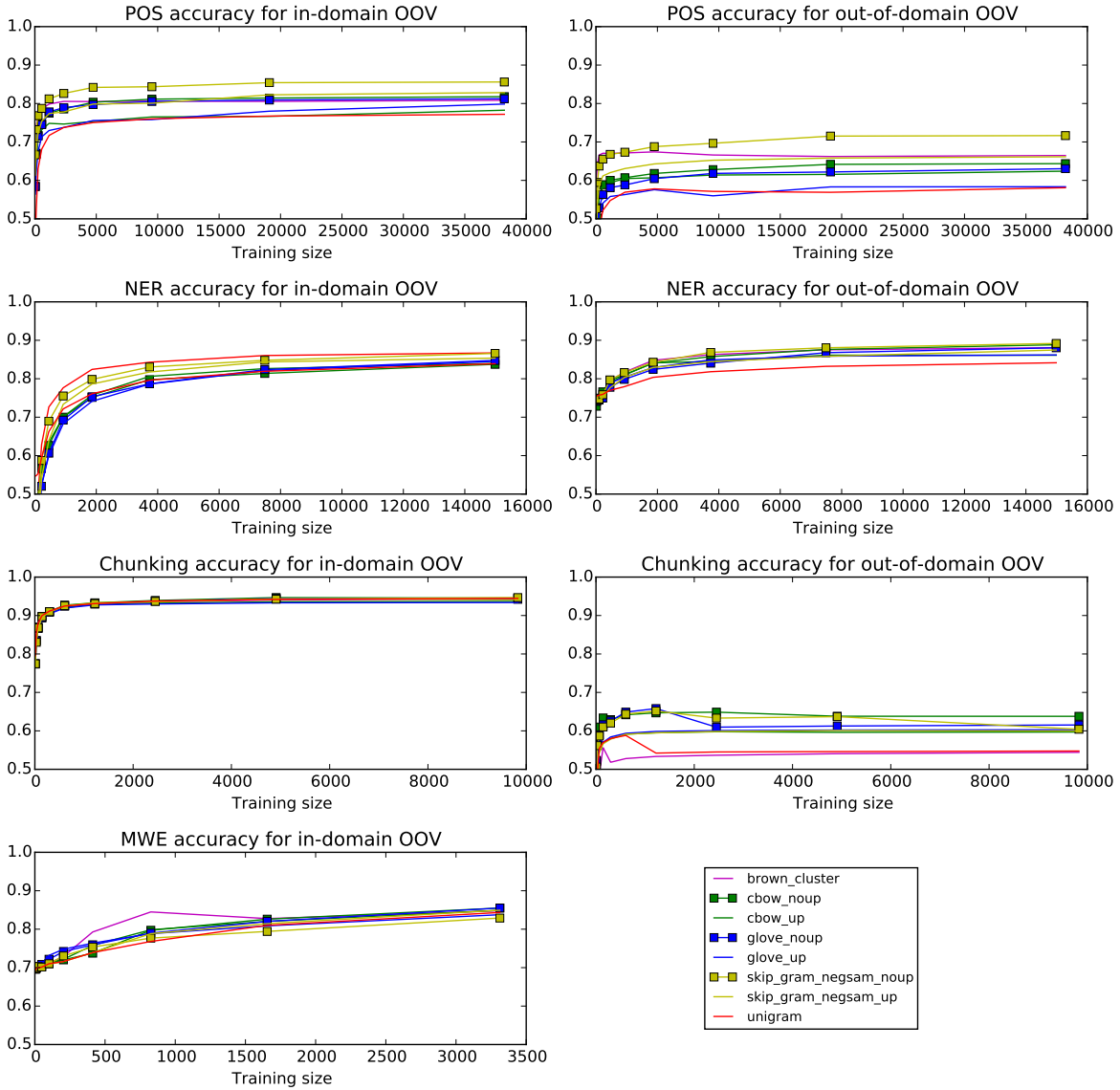


Figure : NER with SKIP-GRAM

RQ4: What is the impact of these word embeddings (with and without updating) on both OOV items (relative to the training data) and out-of-domain data?

RQ5: Are some word embeddings better than others in a sequence labelling context?



Find us!

- github
- wordpress

Thanks!

Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai.
1992.
Class-based n-gram models of natural language.
Computational Linguistics, 18:467–479.

Ciprian Chelba, Tomas Mikolov, Mike Schuster, Qi Ge, Thorsten Brants, Phillipp Koehn, and Tony Robinson.
2013.
One billion word benchmark for measuring progress in statistical language modeling.
Technical report, Google.

Lushan Han, Abhay L. Kashyap, Tim Finin, James Mayfield, and Johnathan Weese.
2013.

UMBC EBIQUITY CORE: Semantic textual similarity systems. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics*, pages 44–52, Atlanta, USA.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014.
GloVe: Global vectors for word representation.
In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 1532–1543, Doha, Qatar.