COMPOSITIONAL MORPHOLOGY FOR WORD REPRESENTATIONS AND LANGUAGE MODELLING

Jan Botha, Phil Blunsom

ICML 2014, Beijing
Motivating Example

What we see

The king finally abdicated after years of unkingly conduct.
Motivating Example

What we see

The king finally abdicated after years of unkingly conduct.

Wait what – unkingly?
Motivating Example

What we see

The king finally abdicated after years of unkingly conduct.

Wait what – unkingly?

unkingly ʌnˈkɪŋli

*a word you have probably never seen, but still understand*
Motivating Example

What we see

The king finally abdicated after years of unkingly conduct.

Wait what – unkingly?

unkingly ʌnˈkɪŋli

a word you have probably never seen, but still understand

⇒ compositional morphology in action
Motivating Example

What we see

The king finally abdicated after years of unkingly conduct.

Wait what – unkingly?

unkingly \(\text{\texttt{\textasciitilde\textasciitilde\textasciitilde\textasciitilde\textasciitilde}}\) a word you have probably never seen, but still understand

⇒ compositional morphology in action

What our models see (mostly)

10 2 95 529 11 88 21 50 74 239
Motivating Example

**What we see**

The *king* finally abdicated after years of *unkingly* conduct.

Wait what – unkingly?

*unkingly* ʌnˈkɪŋli

*a word you have probably never seen, but still understand*

⇒ compositional morphology in action

**What our models see (mostly)**

| 10 | 2  | 95 | 529 | 11 | 88 | 21 | 50 | 74 | 239 |
Motivating Example 2

Other languages display still more variation

Czech Conjugation

Čístit (to clean)
Čístím
Čístiš
Čistí
Čístíme
Čístíte
Čistil
Čištěn
Čisti
Čístěte
Čístěme

Turkish Productive Derivation

Avrupa (Europe)
Avrupalı (of Europe)
Avrupalılaş (become of Europe)
Avrupalılaştırır (to Europeanise)
Avrupalılaştırırama (be unable to Europeanise)
Avrupalılaştırıramadık (we were unable to Europeanise)
...
Motivating Example 2

Other languages display still more variation

Czech Conjugation

ˇcistit (to clean)
ˇcistím
ˇcistíš
ˇcistí
ˇcistíme
ˇcistíte
ˇcistíl
ˇcištěn
ˇcisti
ˇcistěte
ˇcistěme

Turkish Productive Derivation

Avrupa (Europe)
Avrupalı (of Europe)
Avrupalılaş (become of Europe)
Avrupalılaştırır (to Europeanise)
Avrupalılaştırırama (be unable to Europeanise)
Avrupalılaştırıramadık (we were unable to Europeanise)
...

⇒ we should model morphemes!
REPRESENTING WORDS

- Discrete set?
  \[
  \{a, \text{aardvark}, \ldots, \text{account, accounted, accounting}, \ldots\}
  \]
REPRESENTING WORDS

- Discrete set?
  \{a, aardvark, \ldots, account, accounted, accounting, \ldots\}

- Vector space?
<table>
<thead>
<tr>
<th>Motivation</th>
<th>Proposed Method</th>
<th>Experiments</th>
</tr>
</thead>
</table>

**Extract from Collobert & Weston Embeddings**
EXTRACT FROM COLLOBERT & WESTON EMBEDDINGS
EXTRACT FROM COLOBOERT & WESTON EMBEDDINGS
MORPHEME VECTORS

Existing word vectors already capture some morphology.

▶ banks → bank ≈ kings → king ≈ queens → queen

(Mikolov et al. 2013)
MORPHEME VECTORS

Existing word vectors already capture some morphology.

▶ $\mathbf{\text{banks}} - \mathbf{\text{bank}} \approx \mathbf{\text{kings}} - \mathbf{\text{king}} \approx \mathbf{\text{queens}} - \mathbf{\text{queen}}$

(Logical extension:)

▶ $\mathbf{\text{kings}} \approx \mathbf{\text{king}} + \mathbf{\text{-s}}$

▶ $\mathbf{\text{unkingly}} \approx \mathbf{\text{un}}^- + \mathbf{\text{king}} + \mathbf{\text{-ly}}$

(Mikolov et al. 2013)
MORPHEME VECTORS

Existing word vectors already capture some morphology.

▶ banks $\rightarrow$ bank $\approx$ kings $\rightarrow$ king $\approx$ queens $\rightarrow$ queen

(Logical extension:)

▶ kings $\approx$ king $+$ -s

▶ unkingly $\approx$ un$-$ + king $+$ -ly

HOW TO...

▶ obtain morpheme vectors
▶ compose morpheme vectors
▶ do it all *within* a language model usable in an MT decoder
MORPHOLOGICAL COMPOSITION AS ADDITION

Literally, word = sum of its parts?
MORPHOLOGICAL COMPOSITION AS ADDITION

Literally, word = sum of its parts?

Problems:
- bag of morphemes:  \[ \text{hang} + \text{over} \neq \text{over} + \text{hang} \]
- non-compositionality:  \[ \text{greenhouse} \neq \text{green} + \text{house} \]
MORPHOLOGICAL COMPOSITION AS ADDITION

Literally, word = sum of its parts?

Problems:

- bag of morphemes: \[ \text{hang} + \text{over} \neq \text{over} + \text{hang} \]
- non-compositionality: \[ \text{greenhouse} \neq \text{green} + \text{house} \]

PRAGMATIC SOLUTION

include word identity as component too:

\[ \text{greenhouse} \equiv \text{green}_{\text{stem}} + \text{house}_{\text{stem}} \]
\[ \text{unkingly} \equiv \text{un}_{\text{pre}} + \text{king}_{\text{stem}} + \text{ly}_{\text{suf}} \]
MORPHOLOGICAL COMPOSITION AS ADDITION

Literally, word = sum of its parts?

Problems:

- bag of morphemes: $\rightarrow$ hang + over $\neq$ over + hang
- non-compositionality: $\rightarrow$ greenhouse $\neq$ green + house

PRAGMATIC SOLUTION

include word identity as component too:

$\rightarrow$ greenhouse $\equiv$ greenhouse$_{id}$ + green$_{stem}$ + house$_{stem}$

$\rightarrow$ unkingly $\equiv$ unkingly$_{id}$ + un$_{pre}$ + king$_{stem}$ + ly$_{suf}$
SIMPLEST VECTOR-BASED PROBABILISTIC LM

LBL (Log-bilinear model)  (Mnih & Hinton, 2007; Mnih & Teh, 2012)

```
C_1  +  C_2  \implies P(\text{ideas} | \text{colorless, green})
```

“colorless green ideas sleep furiously .”
ADD MORPHEME VECTORS INSIDE LM

LBL++

```
\[ \begin{array}{c}
colorless_{id} + \\
color_{stem} + \\
less_{suf} \\
\downarrow \\
C_1 \\
colorless \\
\end{array} \quad + \quad \begin{array}{c}
green_{id} \\
green_{stem} \\
\downarrow \\
C_2 \\
green \\
\end{array} \quad \Rightarrow \quad \begin{array}{c}
\text{Ideas} \\
\text{Ideas}_{id} + \\
\text{Ideas}_{stem} + \\
\text{Suf}_{suf} \\
\end{array} \quad \Rightarrow \quad P(\text{ideas} \mid \text{colorless, green})
\]
```

“colorless green ideas sleep furiously .”
COMPUTATIONAL EFFICIENCY

Problem:
Each probability query requires normalisation over vocabulary.

- $O(\text{vocab size})$
- rich morphology $\Rightarrow$ large vocabulary
**COMPUTATIONAL EFFICIENCY**

**Problem:**
Each probability query requires normalisation over vocabulary.
- \( O(\text{vocab size}) \)
- rich morphology \( \Rightarrow \) large vocabulary

**Solution:** Decompose model using word classes

\[
P(\text{word} \mid \text{history}) = P(\text{class(word)} \mid \text{history}) \\
	\times P(\text{word} \mid \text{class(word)}, \text{history})
\]

- use unsupervised Brown-clustering
- each LM query becomes \( 2 \times O(\sqrt{\text{vocab size}}) \)

\( \Rightarrow \) fast enough for MT-decoding
**Evaluation Overview**

**Setup**

- 4-gram models
- Czech, English, French, German, Spanish, Russian
- train on 20–50m tokens
- large vocabularies (exclude 5% of singletons)
Evaluation Overview

Setup

- 4-gram models
- Czech, English, French, German, Spanish, Russian
- train on 20–50m tokens
- large vocabularies (exclude 5% of singletons)

Three evaluation contexts:

- Perplexity on test data
- Word similarity rating
- Machine translation
EVALUATION OVERVIEW

Three evaluation contexts:

- Perplexity on test data
- Word similarity rating
- Machine translation
PERPLEXITY IMPROVEMENTS BY LANGUAGE

CLBL→CLBL++

<table>
<thead>
<tr>
<th>Language</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>683→643</td>
</tr>
<tr>
<td>DE</td>
<td>422→404</td>
</tr>
<tr>
<td>EN</td>
<td>281→273</td>
</tr>
<tr>
<td>ES</td>
<td>207→203</td>
</tr>
<tr>
<td>FR</td>
<td>232→227</td>
</tr>
<tr>
<td>RU</td>
<td>313→300</td>
</tr>
</tbody>
</table>
PERPLEXITY IMPROVEMENTS ON GERMAN

CLBL→CLBL++ (BREAK-DOWN BY TOKEN FREQUENCY)
EVALUATION OVERVIEW

Three evaluation contexts:

- Perplexity on test data
- Word similarity rating
- Machine translation
Three evaluation contexts:

- Perplexity on test data
- **Word similarity rating**
- Machine translation
**WORD SIMILARITY RATING**

- CLBL word vectors; unknown test word $\Rightarrow$ generic unknown

![Graph showing Spearman $\rho \times 100$ for different languages and datasets.

- **EN\(_{(RW)}\)**
- **EN\(_{(WS353)}\)**
- **DE\(_{(Gur)}\)**
- **DE\(_{(ZG)}\)**
- **FR\(_{(RG)}\)**
**Word Similarity Rating**

- CLBL word vectors; unknown test word $\Rightarrow$ generic unknown
- CLBL++ composed vectors; unknown test word $\Rightarrow$ generic unknown

![Graph showing Spearman $\rho \times 100$ results for different languages and methods.](image)
WORD SIMILARITY RATING

- CLBL word vectors; unknown test word $\Rightarrow$ generic $\overrightarrow{\text{unknown}}$
- CLBL++ composed vectors; unknown test word $\Rightarrow$ generic $\overrightarrow{\text{unknown}}$
- CLBL++ composed vectors; unknown test word $\Rightarrow\sum\text{known}$$\overrightarrow{\text{morphemes}}$

Bar chart showing Spearman $\rho \times 100$ for different language pairs and vector methods.
**Evaluation Overview**

Three evaluation contexts:

- Perplexity on test data
- **Word similarity rating**
- Machine translation
**Evaluation Overview**

Three evaluation contexts:

- Perplexity on test data
- Word similarity rating
- **Machine translation**
MACHINE TRANSLATION EVALUATION

How to use the LM?

- rescore n-best list < rescore lattice < decoder feature
MACHINE TRANSLATION EVALUATION

How to use the LM?

- rescore n-best list < rescore lattice < decoder feature

Hierarchical-phrase based decoder (cdec)

- Baseline: Kneser-Ney LM feature
- Test: Kneser-Ney LM feature + CLBL feature
Machine Translation Evaluation

How to use the LM?
  ▶ rescore n-best list < rescore lattice < decoder feature

Hierarchical-phrase based decoder (cdec)
  ▶ Baseline: Kneser-Ney LM feature
  ▶ Test: Kneser-Ney LM feature + CLBL feature

CLBL speed-up from:
  ▶ class decomposition
  ▶ cache normalisers on-the-fly
Translation Quality (Bleu) for translating into given language

higher better
TRANSLATION QUALITY (BLEU)
FOR TRANSLATING INTO GIVEN LANGUAGE

higher better
**Translation Quality (BLEU)**

For translating into given language

- **CS**
- **DE**
- **ES**
- **FR**
- **RU**
- **EN**

- Kneser-Ney
- with CLBL
- with CLBL++

*higher better*
QUALITATIVE EVALUATION: ENGLISH AFFIX VECTORS
Simple, scaleable, unsupervised method for integrating morphology into vector-based LM

- improvements in three evaluation settings
- translation with normalised NLM works

Software released shortly
www.clg.ox.ac.uk/resources
{Jan.Botha,Phil.Blunsom}@cs.ox.ac.uk
Simple, scaleable, unsupervised method for integrating morphology into vector-based LM

- improvements in three evaluation settings
- translation with normalised NLM works

Software released shortly

www.clg.ox.ac.uk/resources
{Jan.Botha,Phil.Blunsom}@cs.ox.ac.uk