Monolingual and Cross-Lingual Information Retrieval Models Based on (Bilingual) Word Embeddings

Ivan Vulić and **Marie-Francine Moens**

KU Leuven
Department of Computer Science

ivan.vulic@cs.kuleuven.be marie.francine.moens@cs.kuleuven.be

SIGIR 2015, Santiago de Chile; August 11, 2015

Presentation Plan I

KU LEUVEN

I. Learning Bilingual Word Embeddings (BWEs)

- Dense word representations, word embeddings, bilingual word embeddings
- Monolingual and bilingual embedding spaces
- Multilingual text data $\rightarrow$ why document-aligned data?
- New BWE learning model: BWESG $\rightarrow$ learning monolingual and bilingual embedding spaces

Presentation Plan II

KU LEUVEN

II. BWEs in IR

- Semantically-aware representations in the ad-hoc retrieval process?
- From word representations to query and document representations
- Monolingual embeddings $\rightarrow$ monolingual retrieval; Bilingual embeddings $\rightarrow$ cross-lingual retrieval
- The same conceptual model of retrieval for MoIR and CLIR with bilingual embeddings spaces!
- Results and discussion

Part I: Learning BWEs

Learning Word Representations

KU LEUVEN

Key idea

Distributional hypothesis → words with similar meanings are likely to appear in similar contexts

[Harris, Word 1954]

shouts:

"Meaning as use!"

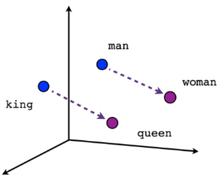
calmly states:

"You shall know a word by the company it keeps."

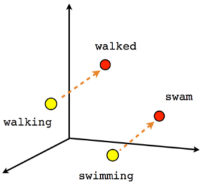
Word Embeddings

Dense representations → real-valued low-dimensional vectors
(seen already? LSI?)

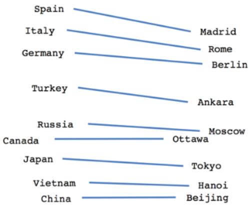
Word embedding induction
→ learn word-level features which generalize well across tasks and languages
→ **bilingual word embeddings** (this talk)
Word embeddings capture interesting and universal features:



Male-Female

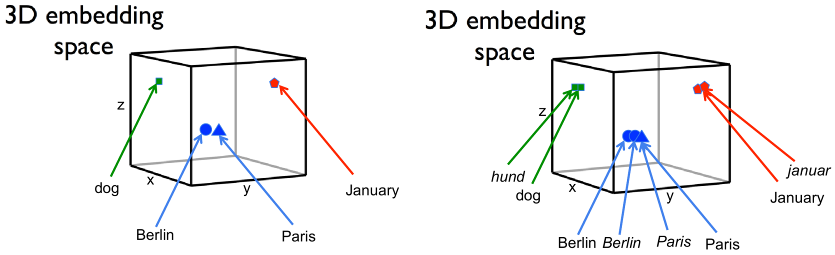


Verb tense



Country-Capital

Embedding Spaces = Semantic Spaces



Monolingual vs. **Bilingual**

[Image courtesy of Stephan Gouws]

Bilingual Word Embeddings

KU LEUVEN

Representation of a word $w_1^S \in V^S$:

$$vec(w_1^S) = [f_1^1, f_2^1, \ldots, f_{dim}^1]$$

Exactly the same representation for $w_2^T \in V^T$:

$$vec(w_2^T) = [f_1^2, f_2^2, \ldots, f_{dim}^2]$$

Language-independent word representations in the same shared semantic (or *embedding*) space!

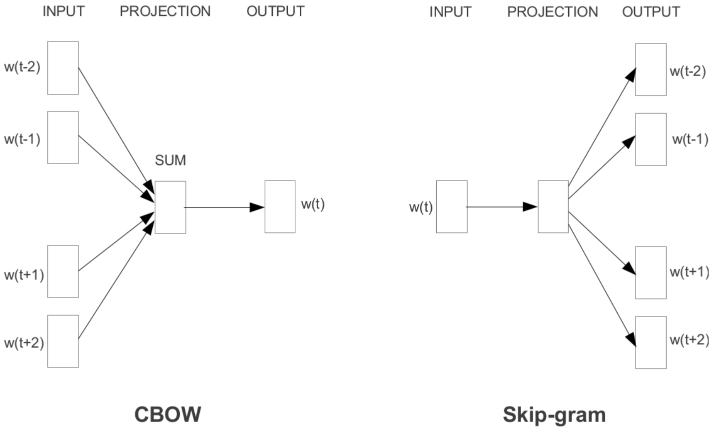
Word representation $\rightarrow$ A dense real-valued dim-dimensional vector, these dimensions are no longer interpretable (unlike with other semantic representations).

Back to Monolingual...

KU LEUVEN

Skip-gram with negative sampling (SGNS)

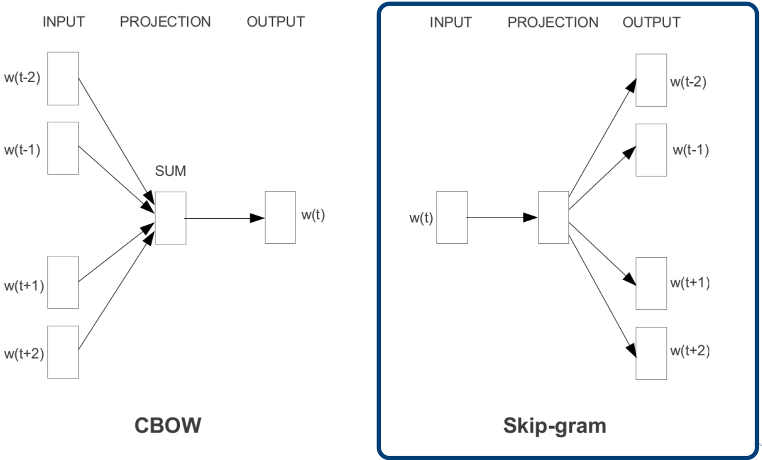
[Mikolov et al.: NIPS 2013]



Back to Monolingual...

Skip-gram with negative sampling (SGNS)

[Mikolov et al.; NIPS 2013]



Back to Monolingual...

KU LEUVEN

Skip-gram with negative sampling (SGNS)
[Mikolov et al.; NIPS 2013]

Learning from the set D of *(word, context)* pairs observed in a corpus:
$(w, v) = (w(t), w(t \pm i))$; $i = 1, ..., cs$; cs = context window size

SG learns to predict the context of the pivot word

John saw a cute gray huhblub running in the field.

D = (huhblub, cute), (huhblub, gray), (huhblub, running), (huhblub, in)
$vec(huhblub) = [-0.23, 0.44, -0.76, 0.33, 0.19, \ldots]$

Back to Monolingual...

Negative sampling = learning using both positive ("observed") examples (set D), and negative ("unobserved") examples (set D')

SGNS is actually doing something very similar to the older approaches $\rightarrow$ factorizing the traditional word-context matrix! [Levy et al., NIPS 2014, TACL 2015]

More research focused on learning monolingual WEs:

- Full-fledged neural-net approaches [Bengio et al., JMLR 2003; Collobert and Weston, ICML 2008]
- Other factorization methods (e.g., Hellinger PCA) [Lebret and Collobert, EACL 2014]
- GloVe [Pennington et al., EMNLP 2014]
- ...

SGNS - A More Formal Summary

KU LEUVEN

Probability for one word-context pair (w, v):

$$P(D = 1 | w, v, \theta) = \frac{1}{1 + \exp(-\vec{w} \cdot \vec{v_c})}$$

General objective:

$$J = \arg\max_{\theta} \sum_{(w,v) \in D} \log \frac{1}{1 + \exp(-\vec{w} \cdot \vec{v_c})}$$

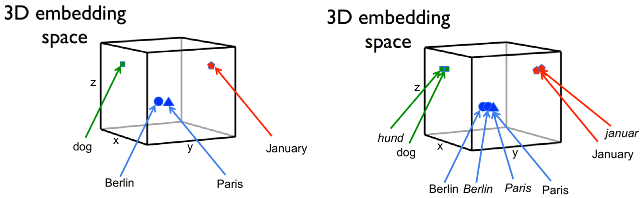
General objective with negative sampling:

$$J = \arg\max_{\theta} \sum_{(w,v) \in D} \log \frac{1}{1 + \exp(-\vec{w} \cdot \vec{v_c})} + \sum_{(w,v') \in D'} \log \frac{1}{1 + \exp(\vec{w} \cdot \vec{v'_c})}$$

[Goldberg and Levy: word2vec explained; arXiv 2014]

And Now Back to Bilingual...

Generalizing the WE learning in bilingual settings using the similar principles...

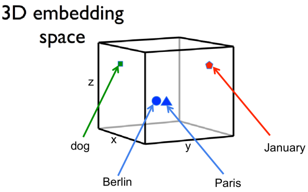


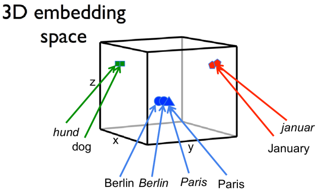
1. Align pretrained monolingual embedding spaces **(offline)** *using dictionaries* [Mikolov et al., arXiv 2013; Lazaridou et al., ACL 2015]
2. Jointly learn and align embeddings **(online)** using *parallel-only data* [Hermann and Blunsom, ACL 2014; Chandar et al., NIPS 2014]
3. Jointly learn and align embeddings **(online)** using *mono* **and** *parallel data* [Gouws et al., ICML 2015; Soyer et al., ICLR 2015, Shi et al., ACL 2015]

And Now Back to Bilingual...

KU LEUVEN

Generalizing the WE learning in bilingual settings using the similar principles...

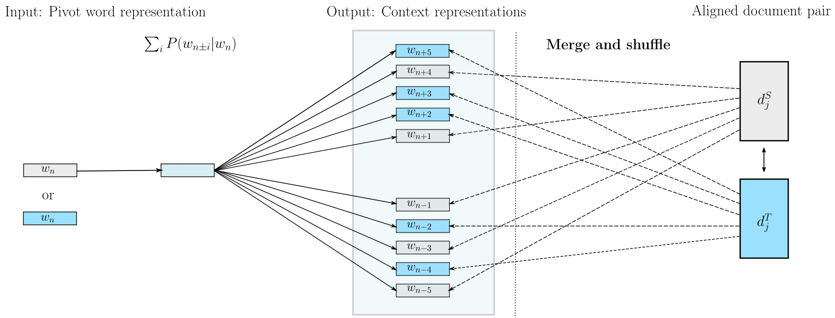




1. Align pretrained monolingual embedding spaces **(offline)** *using dictionaries* [Mikolov et al., arXiv 2013; Lazaridou et al., ACL 2015]
2. Jointly learn and align embeddings **(online)** using *parallel-only data* [Hermann and Blunsom, ACL 2014; Chandar et al., NIPS 2014]
3. Jointly learn and align embeddings **(online)** using *mono* **and** *parallel data* [Gouws et al., ICML 2015; Soyer et al., ICLR 2015, Shi et al., ACL 2015]

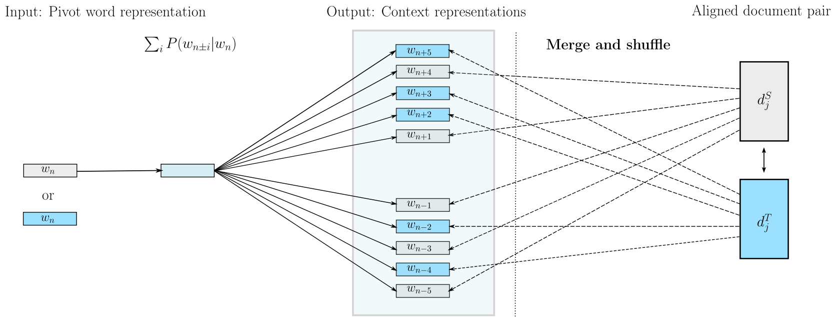
4. Can we do it without readily available dictionaries and parallel data? $\rightarrow$ **Using document-aligned data (e.g., Wikipedia) [our model: BWESG]**

BWEs from Document-Aligned Data



BWEs from Document-Aligned Data

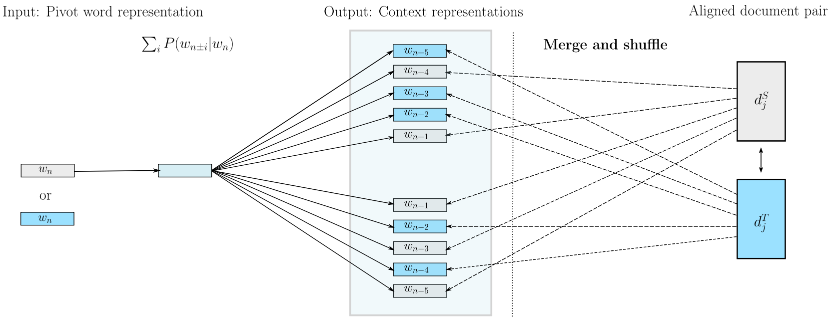
KU LEUVEN



→ **Merge & Shuffle:** Training a SGNS (or any other monolingual model!) on shuffled "pseudo-bilingual" documents →

→ Our model: **BWESG**

BWEs from Document-Aligned Data

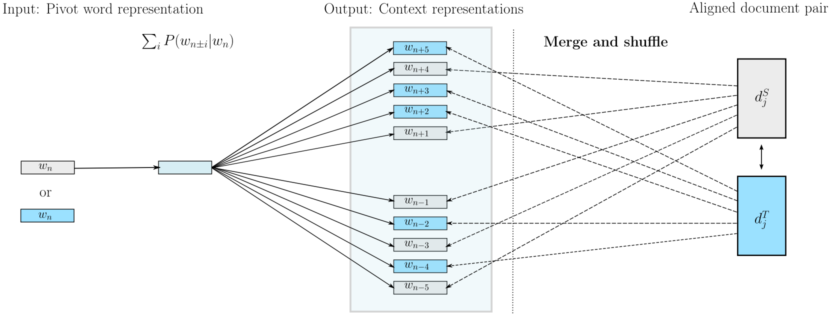


→ **Merge & Shuffle:** Training a SGNS (or any other monolingual model!) on shuffled "pseudo-bilingual" documents →

→ Our model: **BWESG**

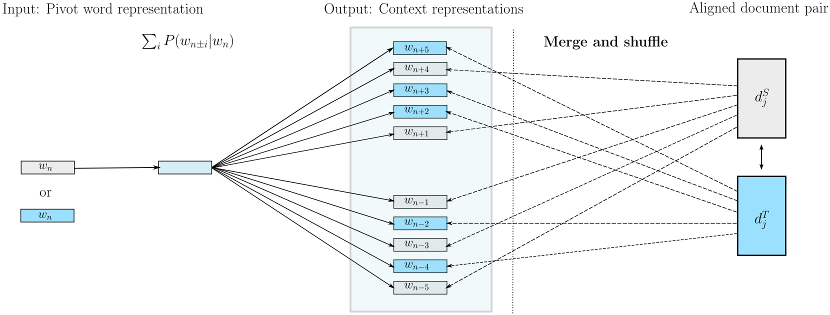
→ **1. dumb shuffling: random (this work)**; *2. slightly more intelligent: length ratio-based (after this work)*; 3. even more intelligent: future work

BWEs from Document-Aligned Data



→ shuffling ensures bilingual (instead of monolingual) contexts → learning a bilingual embedding space jointly **(online)**

BWEs from Document-Aligned Data



→ shuffling ensures bilingual (instead of monolingual) contexts → learning a bilingual embedding space jointly **(online)**

→ **No longer a local model**: Window size controls the number of **document-level** positive samples

BWEs with BWESG - Examples

KU LEUVEN

Spanish-English (ES-EN)			Italian-English (IT-EN)			Dutch-English (NL-EN)		
(1) **reina**	(2) **reina**	(3) **reina**	(1) **madre**	(2) **madre**	(3) **madre**	(1) **schilder**	(2) **schilder**	(3) **schilder**
(Spanish)	(English)	(Combined)	(Italian)	(English)	(Combined)	(Dutch)	(English)	(Combined)
rey	*queen(+)*	*queen(+)*	padre	*mother(+)*	*mother(+)*	kunstschilder	*painter(+)*	*painter(+)*
trono	*heir*	rey	moglie	*father*	padre	schilderij	*painting*	kunstschilder
monarca	*throne*	trono	sorella	*sister*	moglie	kunstenaar	*portrait*	*painting*
heredero	*king*	*heir*	figlia	*wife*	*father*	olieverf	*artist*	schilderij
matrimonio	*royal*	*throne*	figlio	*daughter*	sorella	olieverfschilderij	*canvas*	kunstenaar
hijo	*reign*	monarca	fratello	*son*	figlia	schilderen	*impressionist*	*portrait*
reino	*succession*	heredero	casa	*friend*	figlio	frans	*cubism*	olieverf
reinado	*princess*	*king*	amico	*childhood*	*sister*	nederlands	*art*	olieverfschilderij
regencia	*marriage*	matrimonio	marito	*family*	fratello	componist	*poet*	schilderen
duque	*prince*	*royal*	donna	*cousin*	*wife*	beeldhouwer	*drawing*	*artist*

$$\overrightarrow{reina} - \overrightarrow{woman} + \overrightarrow{man} \approx \overrightarrow{rey}$$

$$\overrightarrow{queen} - \overrightarrow{mujer} + \overrightarrow{hombre} \approx \overrightarrow{king}$$

$$\overrightarrow{reina} - \overrightarrow{mujer} + \overrightarrow{hombre} \approx \overrightarrow{rey}$$

Useful in bilingual lexicon extraction!

Summary of Contributions

KU LEUVEN

- A novel model for learning bilingual word embeddings (BWEs) from non-parallel document-aligned data
- A simple framework for constructing query and document embeddings
- A unified framework for MoIR and CLIR based on (bilingual) word embeddings

Part II: BWEs in IR

Beyond the Level of Words

KU LEUVEN

We learn **word embeddings**:
$vec(\text{huhblub}) = [-0.23, 0.44, -0.76, 0.33, 0.19, \ldots]$
$vec(\text{fluffy}) = [0.31, 0.02, -0.11, -0.28, 0.52, \ldots]$

$\rightarrow$ How to build document and query embeddings?
$vec(\text{huhblup is fluffy}) = ??$

Adapting the framework from *compositional distributional semantics*: [Mitchell and Lapata, ACL 2008; Socher et al., EMNLP 2011; Milajevs et al., EMNLP 2014] and many more...

A generic composition with a **bag-of-words** assumption:
$(d = \{w_1, w_2, \ldots, w_{|N_d|}\})$

$$\overrightarrow{d} = \overrightarrow{w_1} \star \overrightarrow{w_2} \star \ldots \star \overrightarrow{w_{|N_d|}}$$

$\star$ = compositional vector operator (addition, multiplication, tensor product,..)

Document and Query Embeddings

KU LEUVEN

A general framework $\rightarrow$ in this work the simple and effective **additive composition**:

[Mitchell and Lapata, ACL 2008]

$$\overrightarrow{d} = \overrightarrow{w_1} + \overrightarrow{w_2} + \ldots + \overrightarrow{w_{|N_d|}}$$

The dim-dimensional **document embedding** in the same bilingual word embedding space:

$$\overrightarrow{d} = [f_{d,1}, \ldots, f_{d,k}, \ldots, f_{d,dim}]$$

$\rightarrow$ the **ADD-BASIC** composition model

Document and Query Embeddings

KU LEUVEN

A slightly more intelligent idea $\rightarrow$ weighting the summands using their **self information** computed in the target collection:

$$si_w = -\ln \frac{freq(w, \mathcal{DC})}{|N_{\mathcal{DC}}|}$$

$freq(w, \mathcal{DC})$ = frequency of w in the collection

A SI-weighted sum:

$$\overrightarrow{d} = si_{w_1} \cdot \overrightarrow{w_1} + si_{w_2} \cdot \overrightarrow{w_2} + \ldots + si_{w_{|N_d|}} \cdot \overrightarrow{w_{|N_d|}}$$

$\rightarrow$ the **ADD-SI** composition model

Document and Query Embeddings

KU LEUVEN

→ The same principles with **queries**
→ Using only ADD-BASIC

$$\overrightarrow{Q} = \overrightarrow{q_1} + \overrightarrow{q_2} + \ldots + \overrightarrow{q_m}$$

The dim-dimensional **query embedding** in the same bilingual word embedding space:

$$\overrightarrow{Q} = [f_{Q,1}, \ldots, f_{Q,k}, \ldots, f_{Q,dim}]$$

Final BWE-Based MoIR and CLIR I

1. **Induce** a bilingual word embedding space using any BWE induction model $\rightarrow$ *in this work: BWESG*

2. Given is a target document collection $\mathcal{DC} = \{d'_1, \ldots, d'_{N'}\}$. **Compute** dim-dimensional document embeddings $\overrightarrow{d'}$ for each $d' \in \mathcal{DC}$ using the dim-dimensional WEs from the set $\mathcal{BWE}$ obtained in the previous step and a semantic composition model (ADD-BASIC or ADD-SI something anything else).

3. After the query $Q = \{q_1, \ldots, q_m\}$ is issued in language L_S, **compute** a dim-dimensional query embedding using the ADD-BASIC composition model.

Final BWE-Based MoIR and CLIR II

4. **For each** $d' \in \mathcal{DC}$, **compute** the semantic similarity score $sim(d', Q)$ which quantifies each document's relevance to the query Q:

$$sim(d', Q) = SF(d', Q) = \frac{\overrightarrow{d'} \cdot \overrightarrow{Q}}{|\overrightarrow{d'}| \cdot |\overrightarrow{Q}|}$$

5. **Rank** all documents from $\mathcal{DC}$ according to their similarity scores from the previous step.

WE-VS: WE-based MoIR and CLIR models (using ADD-BASIC)

Part IIb: Experiments

BWESG Training Setup

KU LEUVEN

- Stochastic gradient descent with a default global learning rate 0.025
- Other default `word2vec` parameters: subsampling rate $1e-4$, negative sampling with 25 negative samples, 15 epochs
- 10 random corpora shuffles, although **we advocate the use of a more intelligent shuffling procedure** (developed after the paper was released)
- $d = 100 - 800$ in steps of 100
- $cs = 10 - 100$ in steps of 10

Training Collections

KU LEUVEN

[English|Dutch] $\rightarrow$ **[English|Dutch]** retrieval

Exactly the same setup as in: [Vulić et al., Information Retrieval 2013, ECIR 2013]

Training data	Europarl	$6,206$ documents (parallel)
	Wikipedia	$7,612$ documents (comparable)
Vocabulary size	English	$76,555$ words
	Dutch	$71,168$ words

$\rightarrow$ Stop words removed
$\rightarrow$ We exploit document-level alignments as the only bilingual signal (even for Europarl)

Test Collections

[English|Dutch] $\rightarrow$ **[English|Dutch]** retrieval (using CLEF 2001-2003 campaigns)

Monolingual				
Direction	$\mathcal{DC}$	# Docs	Query Set	# Queries
EN$\rightarrow$EN 2001	LAT	$110,861$	EN'01: 41-90	47
EN$\rightarrow$EN 2002	LAT	$110,861$	EN'02: 91-140	42
EN$\rightarrow$EN 2003	LAT+GH	$166,753$	EN'03: 141-200	53
NL$\rightarrow$NL 2001	NC+AD	$190,604$	NL'01: 41-90	50
NL$\rightarrow$NL 2002	NC+AD	$190,604$	NL'02: 91-140	50
NL$\rightarrow$NL 2003	NC+AD	$190,604$	NL'03: 141-200	56

Test Collections

KU LEUVEN

[English|Dutch] → **[English|Dutch]** retrieval (using CLEF 2001-2003 campaigns)

Cross-lingual				
Direction	$\mathcal{DC}$	# Docs	Query Set	# Queries
NL→EN 2001	LAT	$110,861$	NL'01: 41-90	47
NL→EN 2002	LAT	$110,861$	NL'01: 91-140	42
NL→EN 2003	LAT+GH	$166,753$	NL'03: 141-200	53
EN→NL 2001	NC+AD	$190,604$	EN'01: 41-90	50
EN→NL 2002	NC+AD	$190,604$	EN'02: 91-140	50
EN→NL 2003	NC+AD	$190,604$	EN'03: 141-200	56

→ Queries extracted from the *title* + *description* fields

→ Stop words removed → Measuring MAP

Models in Comparison

KU LEUVEN

Single models:

1. **WE-VS**: Our WE-based retrieval model

2. **LM-UNI**: Unigram query likelihood language model with standard Dirichlet smoothing

3. **LDA-IR**: Semantically-aware (Bi)LDA-based QL model [Wei and Croft, SIGIR 2006; Vulić et al, IR 2013]

A detailed description of all the models along with their parameter setup in the paper!

Models in Comparison

Combined models:

1. **LM-UNI+LDA-IR**: A linear combination of the two single models:
[Wei and Croft, SIGIR 2006; Vulić et al, IR 2013]

$$P(q_i|d) = \lambda P_{lda}(q_i|d) + (1-\lambda)P_{lm}(q_i|d)$$

2. **LM-UNI+WE-VS**: A linear combination of LM-UNI and WE-VS (to directly compare the "quality of semantic awareness" in the retrieval process)

x. **GT+LM+LDA** (only for CLIR): Translating a query using *Google Translate*, and then employing LM-UNI+LDA-IR on the translated query

Again, a detailed description of all the models along with their parameter setup in the paper!

Results - MoIR

KU LEUVEN

Model	EN→EN 2001	EN→EN 2002	EN→EN 2003	NL→NL 2001	NL→NL 2002	NL→NL 2003
LM-UNI	.381	.360	.359	.256	.323	.357
LDA-IR	.279	.216	.241	.131	.143	.130
dim:300; *cs*:60						
WE-VS	.324x	.258x	.257y	.203x	.237x	.224x
dim:600; *cs*:60						
WE-VS	.329x	.281x	.262y	.204x	.262x	.231x
LM+LDA	.399	.360	.379	.260	.326	.357
dim:300; *cs*:60						
LM+WE (λ=0.3)	.412y	.381x	.401y	.271x	.349x	.372x
LM+WE (λ=0.5)	.429x	.394x	.407x	.279x	.370x	.382x
LM+WE (λ=0.7)	.451x	.392y	.389	.270	.364x	.373y
dim:600; *cs*:60						
LM+WE (λ=0.3)	.419y	.382x	.403y	.274x	.350x	.373x
LM+WE (λ=0.5)	.436x	.391x	.408x	.282x	.371x	.383x
LM+WE (λ=0.7)	.430x	.392y	.381	.268	.367x	.374y

Results - MoIR

	EN→EN			NL→NL		
Model	2001	2002	2003	2001	2002	2003
LM-UNI	.381	.360	.359	.256	.323	.357
LDA-IR	.279	.216	.241	.131	.143	.130
dim:300; cs:60						
WE-VS	.324x	.258x	.257y	.203x	.237x	.224x
dim:600; cs:60						
WE-VS	.329x	.281x	.262y	.204x	.262x	.231x
LM+LDA	.399	.360	.379	.260	.326	.357
dim:300; cs:60						
LM+WE (λ=0.3)	.412y	.381x	.401y	.271x	.349x	.372x
LM+WE (λ=0.5)	.429x	.394x	.407x	.279x	.370x	.382x
LM+WE (λ=0.7)	.451x	.392y	.389	.270	.364x	.373y
dim:600; cs:60						
LM+WE (λ=0.3)	.419y	.382x	.403y	.274x	.350x	.373x
LM+WE (λ=0.5)	.436x	.391x	.408x	.282x	.371x	.383x
LM+WE (λ=0.7)	.430x	.392y	.381	.268	.367x	.374y

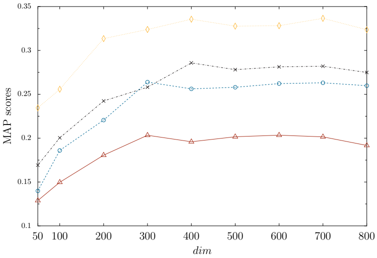
Results - MoIR

Model	EN→EN 2001	EN→EN 2002	EN→EN 2003	NL→NL 2001	NL→NL 2002	NL→NL 2003
LM-UNI	.381	.360	.359	.256	.323	.357
LDA-IR	.279	.216	.241	.131	.143	.130
dim:300; cs:60						
WE-VS	.324x	.258x	.257y	.203x	.237x	.224x
dim:600; cs:60						
WE-VS	.329x	.281x	.262y	.204x	.262x	.231x
LM+LDA	.399	.360	.379	.260	.326	.357
dim:300; cs:60						
LM+WE (λ=0.3)	.412y	.381x	.401y	.271x	.349x	.372x
LM+WE (λ=0.5)	.429x	.394x	.407x	.279x	.370x	.382x
LM+WE (λ=0.7)	.451x	.392y	.389	.270	.364x	.373y
dim:600; cs:60						
LM+WE (λ=0.3)	.419y	.382x	.403y	.274x	.350x	.373x
LM+WE (λ=0.5)	.436x	.391x	.408x	.282x	.371x	.383x
LM+WE (λ=0.7)	.430x	.392y	.381	.268	.367x	.374y

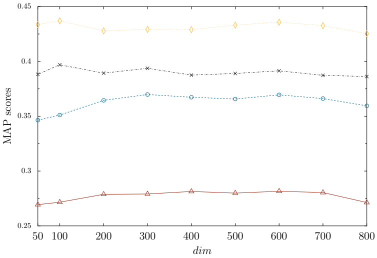
Results - MoIR II

KU LEUVEN

Testing the influence of dimensionality...



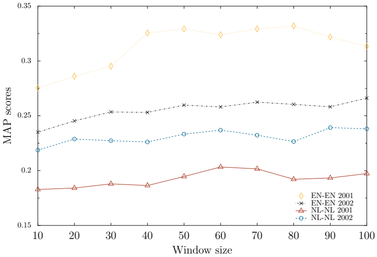
(a) WE-VS, $cs = 60$



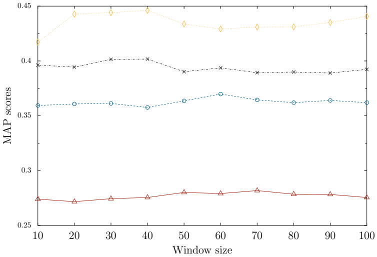
(b) LM-UNI+WE-VS, $cs = 60$

Results - MoIR III

..and window size... (controlling the data dropout)



(c) WE-VS, $dim = 300$



(d) LM-UNI+WE-VS, $dim = 300$

Results - CLIR

KU LEUVEN

	NL→EN			EN→NL		
Model	2001	2002	2003	2001	2002	2003
LM-UNI	.094	.108	.092	.078	.125	.112
LDA-IR	.197	.139	.123	.145	.137	.171
dim:300; cs:60						
WE-VS	.187	.204x	.120	.174	.185y	.157
dim:600; cs:60						
WE-VS	.222y	.230x	.127	.178y	.219x	.181
LM+LDA	.267	.225	.199	.225	.268	.278
GT+LM+LDA	.307	.275	.248	.230	.240	.244
dim:300; cs:60						
LM+WE (λ=0.3)	.189	.273	.197	.101	.159	.150
LM+WE (λ=0.5)	.218	.283y	.220	.113	.184	.167
LM+WE (λ=0.7)	.255	.307x	.219	.180	.209	.208
dim:600; cs:60						
LM+WE (λ=0.3)	.205	.281y	.198	.107	.167	.154
LM+WE (λ=0.5)	.236	.299x	.215	.123	.203	.183
LM+WE (λ=0.7)	.286	.317x	.222	.190	.249	.225

Results - CLIR

Model	NL→EN 2001	NL→EN 2002	NL→EN 2003	EN→NL 2001	EN→NL 2002	EN→NL 2003
LM-UNI	.094	.108	.092	.078	.125	.112
LDA-IR	.197	.139	.123	.145	.137	.171
dim:300; cs:60						
WE-VS	.187	.204x	.120	.174	.185y	.157
dim:600; cs:60						
WE-VS	.222y	.230x	.127	.178y	.219x	.181
LM+LDA	.267	.225	.199	.225	.268	.278
GT+LM+LDA	.307	.275	.248	.230	.240	.244
dim:300; cs:60						
LM+WE (λ=0.3)	.189	.273	.197	.101	.159	.150
LM+WE (λ=0.5)	.218	.283y	.220	.113	.184	.167
LM+WE (λ=0.7)	.255	.307x	.219	.180	.209	.208
dim:600; cs:60						
LM+WE (λ=0.3)	.205	.281y	.198	.107	.167	.154
LM+WE (λ=0.5)	.236	.299x	.215	.123	.203	.183
LM+WE (λ=0.7)	.286	.317x	.222	.190	.249	.225

Results - CLIR

KU LEUVEN

Model	NL→EN 2001	NL→EN 2002	NL→EN 2003	EN→NL 2001	EN→NL 2002	EN→NL 2003
LM-UNI	.094	.108	.092	.078	.125	.112
LDA-IR	.197	.139	.123	.145	.137	.171
dim:300; *cs*:60						
WE-VS	.187	.204x	.120	.174	.185y	.157
dim:600; *cs*:60						
WE-VS	.222y	.230x	.127	.178y	.219x	.181
LM+LDA	.267	.225	.199	.225	.268	.278
GT+LM+LDA	.307	.275	.248	.230	.240	.244
dim:300; *cs*:60						
LM+WE (λ=0.3)	.189	.273	.197	.101	.159	.150
LM+WE (λ=0.5)	.218	.283y	.220	.113	.184	.167
LM+WE (λ=0.7)	.255	.307x	.219	.180	.209	.208
dim:600; *cs*:60						
LM+WE (λ=0.3)	.205	.281y	.198	.107	.167	.154
LM+WE (λ=0.5)	.236	.299x	.215	.123	.203	.183
LM+WE (λ=0.7)	.286	.317x	.222	.190	.249	.225

Results - CLIR

Model	NL→EN 2001	NL→EN 2002	NL→EN 2003	EN→NL 2001	EN→NL 2002	EN→NL 2003
LM-UNI	.094	.108	.092	.078	.125	.112
LDA-IR	.197	.139	.123	.145	.137	.171
dim:300; cs:60						
WE-VS	.187	.204x	.120	.174	.185y	.157
dim:600; cs:60						
WE-VS	.222y	.230x	.127	.178y	.219x	.181
LM+LDA	.267	.225	.199	.225	.268	.278
GT+LM+LDA	.307	.275	.248	.230	.240	.244
dim:300; cs:60						
LM+WE (λ=0.3)	.189	.273	.197	.101	.159	.150
LM+WE (λ=0.5)	.218	.283y	.220	.113	.184	.167
LM+WE (λ=0.7)	.255	.307x	.219	.180	.209	.208
dim:600; cs:60						
LM+WE (λ=0.3)	.205	.281y	.198	.107	.167	.154
LM+WE (λ=0.5)	.236	.299x	.215	.123	.203	.183
LM+WE (λ=0.7)	.286	.317x	.222	.190	.249	.225

Results - CLIR II

	NL→EN			EN→NL		
Model	2001	2002	2003	2001	2002	2003
LM+LDA	.267	.225	.199	.225	.268	.278
GT+LM+LDA	.307	.275	.248	.230	.240	.244
dim:600; cs:60						
LM+WE (λ=0.3)	.205	.281y	.198	.107	.167	.154
LM+WE (λ=0.5)	.236	.299x	.215	.123	.203	.183
LM+WE (λ=0.7)	.286	.317x	.222	.190	.249	.225
dim:600; cs:60						
LM+LDA+WE (λ=0.3)	.277	.263	.210	.229	.288	.283
LM+LDA+WE (λ=0.5)	.281y	.281y	.214	.240	.297y	.290
LM+LDA+WE (λ=0.7)	.302x	.302x	.227	.244y	.311x	.302y

Results - Composition

	Monolingual					
	EN→EN			NL→NL		
Composition	2001	2002	2003	2001	2002	2003
ADD-BASIC (300-60)	.324	.258	.257	.203	.237	.224
ADD-SI (300-60)	.338	.278y	.255	.212	.253y	.227
ADD-BASIC (600-60)	.329	.281	.262	.204	.262	.231
ADD-SI (600-60)	.344y	.301y	.263	.215	.275y	.234

Results - Composition II

KU LEUVEN

	Cross-lingual					
	NL→EN			EN→NL		
Composition	2001	2002	2003	2001	2002	2003
ADD-BASIC (300-60)	.187	.204	.120	.174	.185	.157
ADD-SI (300-60)	.216x	.213y	.122	.189y	.208x	.161
ADD-BASIC (600-60)	.221	.230	.127	.178	.219	.181
ADD-SI (600-60)	.237y	.233	.130	.189	.229x	.184

Summary of Contributions (Repeated)

KU LEUVEN

- A novel model for learning bilingual word embeddings (BWEs) from non-parallel document-aligned data
- A simple framework for constructing query and document embeddings
- A unified framework for MoIR and CLIR based on (bilingual) word embeddings

So, what's next? I

The proposed framework is very general:

Designing other shuffling procedures for BWESG

Building new BWE induction models for the same multilingual data type (remove the need for pseudo-bilingual documents?

Experimenting with other monolingual WE induction models for BWESG besides SGNS

Investigating other BWE induction models (different bilingual signals) in the same (CL)IR pipeline

So, what's next? II

Investigating more elaborate composition models to construct document and query embeddings (what about syntax?)

Testing true *paragraph and phrase embeddings* in the same (CL)IR pipeline

[Le and Mikolov, ICML 2014; Soyer et al., ICLR 2015]

Other (more distant) language pairs, other queries+test collections

Combining the semantic BWE-based knowledge with other IR modeling paradigms (besides the ones mentioned here)

Questions?

Спасибо
Gracias
شكرا
Obrigado
Спасибо
Dank U
Grazie
Ευχαριστώ
Danke
Ευχαριστώ
Merci
Thank You
Ngiyabonga
Dank U
Diolch
Thank You
Ngiyabonga
Tack
Obrigado
Merci
Ευχαριστώ
Dzięki
Danke
Grazie
תודה
Dank U
Terima Kasih
Diolch
Grazie
Merci
Dzięki
Diolch
Gracias
Merci
תודה
Tack
Tack