



Bar-Ilan University

context2vec: Learning Generic Context Embedding with Bidirectional LSTM

Oren Melamud, Jacob Goldberger, Ido Dagan

CoNLL, 2016

What context is

They robbed the _bank_ last night.

- Target: bank
- Sentential context: They robbed the _____ last night.

What context representations are used for

What context representations are used for

- Sentence completion

IBM _____ this company for 100 million dollars.

What context representations are used for

- Sentence completion

IBM _____ this company for 100 million dollars.

- Word sense disambiguation

They robbed the _bank_ last night.

What context representations are used for

- Sentence completion

IBM _____ this company for 100 million dollars.

- Word sense disambiguation

They robbed the _bank_ last night.

- Named entity recognition

I can't find _April_.

What context representations are used for

- Sentence completion

IBM ____ this company for 100 million dollars.

- Word sense disambiguation

They robbed the _bank_ last night.

- Named entity recognition

I can't find _April_.

- More: supersense tagging, coreference resolution, ...

What we want from context representations

- Information on the target slot/word

What we want from context representations

- Information on the target slot/word
- Contextual information \neq sum of context words

What we want from context representations

- Information on the target slot/word
- Contextual information \neq sum of context words

Similar context words, different contextual information

- IBM _____ this company for 100 million dollars.
- IBM bought this company for ___ million dollars.

What we want from context representations

- Information on the target slot/word
- Contextual information \neq sum of context words

Similar context words, different contextual information

- IBM _____ this company for 100 million dollars.
- IBM bought this company for ___ million dollars.

Different context words, similar contextual information

- IBM _____ this company for 100 million dollars.
- I _____ this necklace for my wife's birthday.

What we want from context representations

- Information on the target slot/word
- Contextual information \neq sum of context words

Similar context words, different contextual information

- IBM _____ this company for 100 million dollars.
- IBM bought this company for ___ million dollars.

Different context words, similar contextual information

- IBM _____ this company for 100 million dollars.
- I _____ this necklace for my wife's birthday.

- Context representation \neq sentence representation

- Our goal
 - Sentential context representations
 - More value than sum of words
 - Unsupervised generic learning setting

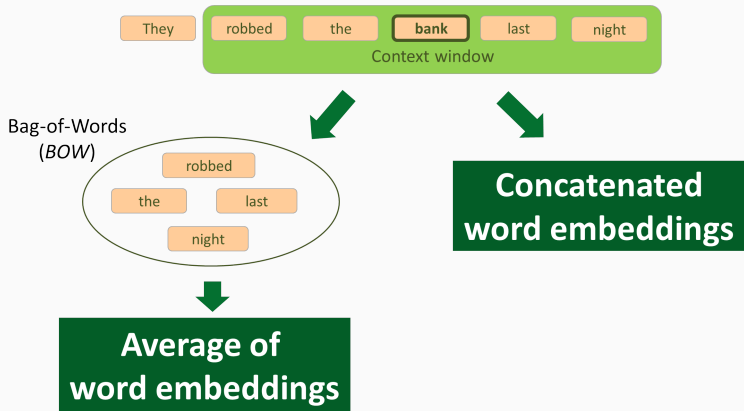
- Our goal
 - Sentential context representations
 - More value than sum of words
 - Unsupervised generic learning setting
- Our model
 - $\text{context2vec} = \text{word2vec} - \text{CBOW} + \text{biLSTM}$

- Our goal
 - Sentential context representations
 - More value than sum of words
 - Unsupervised generic learning setting
- Our model
 - $\text{context2vec} = \text{word2vec} - \text{CBOW} + \text{biLSTM}$
- We show
 - $\text{context2vec} \gg \text{average of word embeddings}$
 - $\text{context2vec} \sim \text{state-of-the-art (more complex models)}$

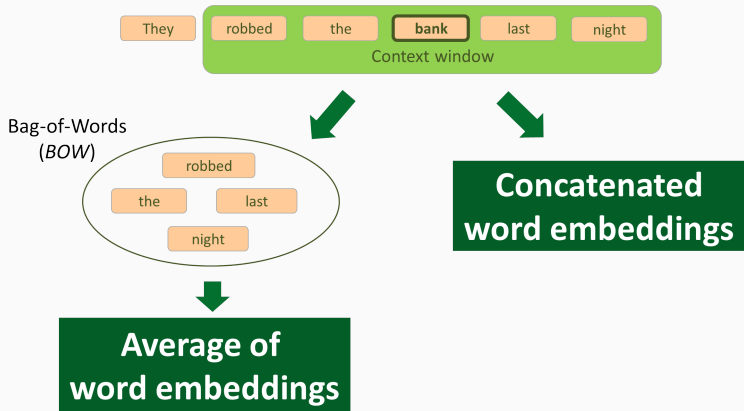
- Our goal
 - Sentential context representations
 - More value than sum of words
 - Unsupervised generic learning setting
- Our model
 - $\text{context2vec} = \text{word2vec} - \text{CBOW} + \text{biLSTM}$
- We show
 - $\text{context2vec} \gg \text{average of word embeddings}$
 - $\text{context2vec} \sim \text{state-of-the-art (more complex models)}$
- Toolkit available for your NLP application

Background

Popular recent context representations

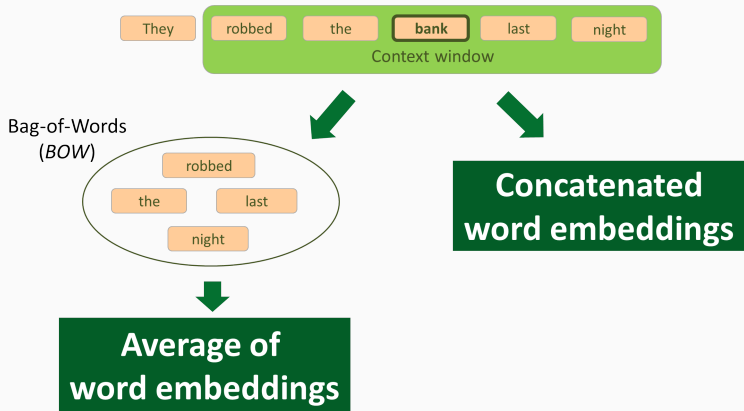


Popular recent context representations



loses word order

Popular recent context representations

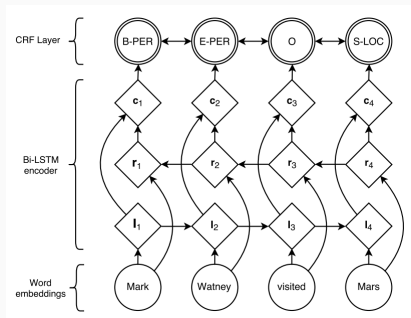


loses word order

Limited scope
Variable-size

Supervised biLSTM with pre-trained word embeddings

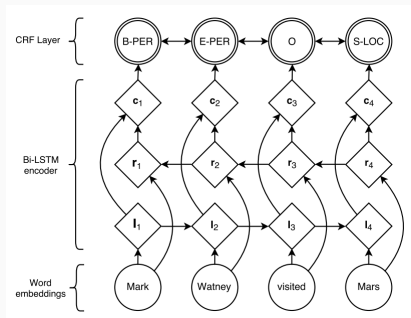
- Word order captured with biLSTM
- Task-specific training
- Supervision is limited in size



NER (Lample, 2016)

Supervised biLSTM with pre-trained word embeddings

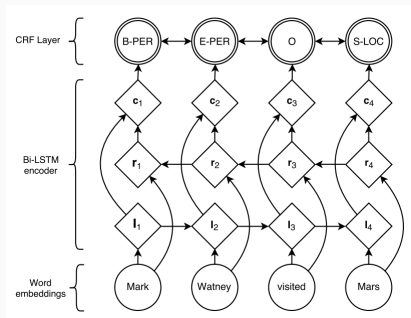
- Word order captured with biLSTM
- Task-specific training
- Supervision is limited in size
- Pre-trained word embeddings carry valuable information from large corpora



NER (Lample, 2016)

Supervised biLSTM with pre-trained word embeddings

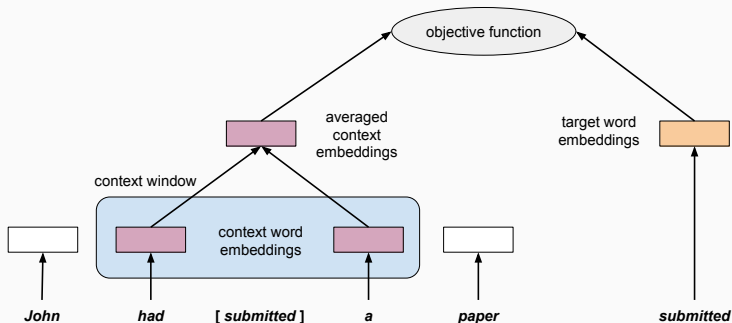
- Word order captured with biLSTM
- Task-specific training
- Supervision is limited in size
- Pre-trained word embeddings carry valuable information from large corpora
- Can we bring even more information?



NER (Lample, 2016)

Model

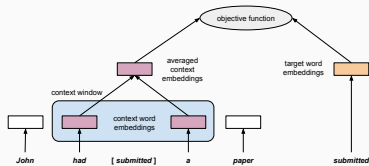
Baseline architecture: word2vec with CBOW



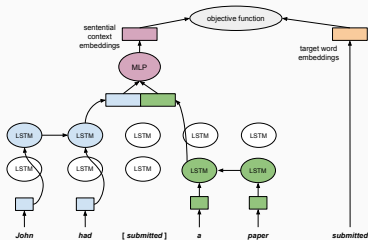
$$S = \sum_{(t,c) \in \text{PAIRS}} \left(\log \sigma(\vec{c}_{avg} \cdot \vec{t}) + \sum_{t' \in \text{NEGS}_{(t,c)}} \log \sigma(-\vec{c}_{avg} \cdot \vec{t}') \right)$$

context2vec = word2vec - CBOW + biLSTM

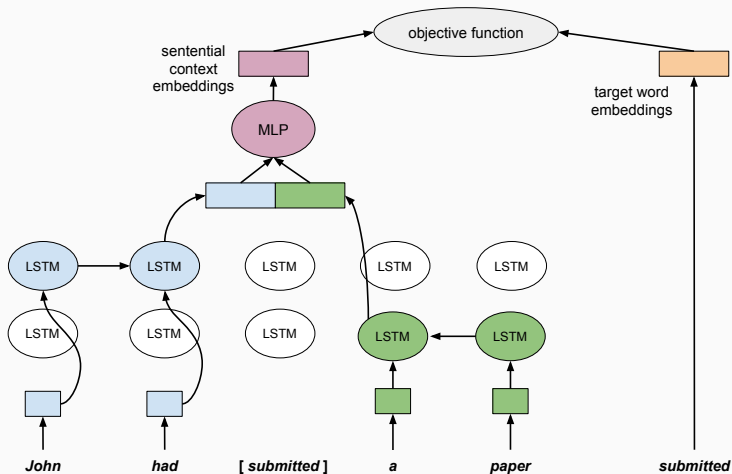
word2vec CBOW



context2vec

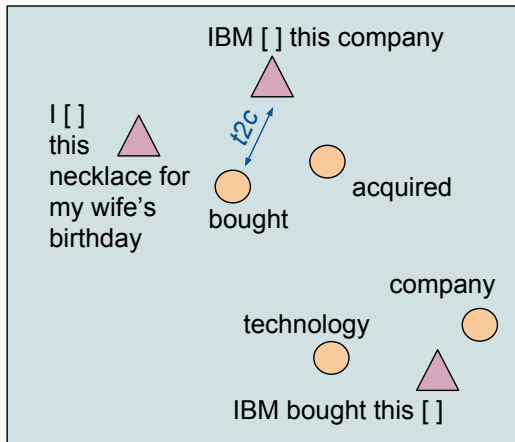


Learning architecture: context2vec




$$S = \sum_{(t,c) \in \text{PAIRS}} \left(\log \sigma(\vec{c}_{c2v} \cdot \vec{t}) + \sum_{t' \in \text{NEGS}_{(t,c)}} \log \sigma(-\vec{c}_{c2v} \cdot \vec{t}') \right)$$

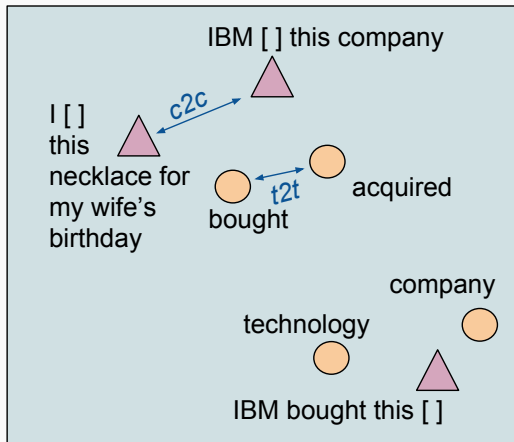
The context2vec embedding space




 target word

 sentential context

The context2vec embedding space



 target word

 sentential context

Evaluation & Results

Evaluation goals

- Standalone evaluation of context2vec
- Using simple cosine similarity measures

Tasks: Sentence completion

I have seen it on him, and could _____ to it.

write

migrate

climb

swear

contribute

- Implementation: Shortest target-context cosine distance
- Benchmark: Microsoft sentence completion challenge (Zweig and Burges, 2011)

Tasks: Lexical substitution

Charlie is a bright boy.

skilled

luminous

vivid

hopeful

smart

- Implementation: Rank by target-context cosine distance
- Benchmarks:
 - Lexical sample (McCarthy and Navigli and Burges, 2007)
 - All-words (Kremer et al., 2014)

Tasks: Supervised word sense disambiguation

This adds a wider
perspective.

Tasks: Supervised word sense disambiguation

TEST

This adds a wider
perspective.

TRAIN

- They add (s2) a touch of humor.
- The minister added (s4) : the process remains fragile.

- Implementation: Shortest context-context cosine distance (kNN)

Tasks: Supervised word sense disambiguation

TEST

This adds a wider perspective.

- Implementation: Shortest context-context cosine distance (kNN)
- Benchmark: Senseval-3 English lexical sample (Mihalcea et al. , 2004)

TRAIN

- They add (s2) a touch of humor.
- The minister added (s4) : the process remains fragile.

	<i>c2v</i>	Avg*
Sentence completion	65.1	49.7
LexSub (sample)	56.0	42.5
LexSub (all-words)	47.9	38.9
WSD	72.8	61.4

* Avg baseline:

- Based on standard Skip-gram word embeddings
- Hyperparameters optimized: context size, word weights.

Results

	<i>c2v</i>	SOTA
Sentence completion	65.1	58.9*
LexSub (sample)	56.0	55.1
LexSub (all-words)	47.9	50.2
WSD	72.8	74.1

- Better recent result 69.2 (Tran, 2016)

Note: SOTA models are generally more complex than *c2v*

Conclusions

- $\text{context2vec} = \text{word2vec} - \text{CBOW} + \text{biLSTM}$

Summary

- $\text{context2vec} = \text{word2vec} - \text{CBOW} + \text{biLSTM}$
- $\text{context2vec} \gg \text{average-of-word-embeddings}$

Summary

- $\text{context2vec} = \text{word2vec} - \text{CBOW} + \text{biLSTM}$
- $\text{context2vec} \gg \text{average-of-word-embeddings}$
- $\text{context2vec} \sim \text{SOTA}$

- $\text{context2vec} = \text{word2vec} - \text{CBOW} + \text{biLSTM}$
- $\text{context2vec} \gg \text{average-of-word-embeddings}$
- $\text{context2vec} \sim \text{SOTA}$

Appealing alternative for
generic context representation

Try context2vec yourself

- The context2vec python toolkit is available at:
<https://github.com/orenmel/context2vec>

Try context2vec yourself

- The context2vec python toolkit is available at:
<https://github.com/orenmel/context2vec>
- Integrate into your NLP application
 - With our pre-trained models
 - Learn your own (choose corpus, dimensionality, etc.)

Try context2vec yourself

- The context2vec python toolkit is available at:
<https://github.com/orenmel/context2vec>
- Integrate into your NLP application
 - With our pre-trained models
 - Learn your own (choose corpus, dimensionality, etc.)
- Potentially more effective or complementary to using pre-trained word embeddings

Try context2vec yourself

- The context2vec python toolkit is available at:
<https://github.com/orenmel/context2vec>
- Integrate into your NLP application
 - With our pre-trained models
 - Learn your own (choose corpus, dimensionality, etc.)
- Potentially more effective or complementary to using pre-trained word embeddings

THANK YOU!

Backup slides

Qualitative example

Sentential Context	Closest target words
This [] is due	item, fact-sheet, offer, pack, card
This [] is due not just to mere luck	offer, suggestion, announcement, item, prize
This [] is due not just to mere luck, but to outstanding work and dedication	award, prize, turnabout, offer, gift
[] is due not just to mere luck, but to outstanding work and dedication	it, success, this, victory, prize-money

Target-context example: bias towards rare words

α	John was [] last year
0.25	born, late, married, out, back
0.50	born, back, married, released, elected
0.75	born, interviewed, re-elected
1.00	starstruck, goal-less, unwed

- α - negative sampling hyperparameter
- Larger α values yield bias towards more rare target words

Context-context example

Query	Furthermore our work in Uganda and Romania [adds] a wider perspective.
<i>context2vec</i> closest	Themes in art have a fascination, since they [add] a subject interest ...
	Richard is joining us every month to pass on tips, and [add] a touch of humour too.
Average closest	The foreign ministers said reforms in Poland and Hungary had made considerable progress but [added] : ...
	Germany had announced the solution [adding] that it hoped Bonn in future ...

- Sentences from Senseval-3 (shortened for readability).