Learning Embeddings for Transitive Verb Disambiguation by Implicit Tensor Factorization

Kazuma Hashimoto
Yoshimasa Tsuruoka

University of Tokyo
Composition: Words $\rightarrow$ Phrases

- Composition models
  - Word embeddings $\rightarrow$ phrase embeddings
- Transitive verbs are good test beds
  - Interaction with their arguments is important!
    - i.e., transitive verb sense disambiguation

31/07/2015 CVSC2015 in Beijing, China
Embeddings of Transitive Verb Phrases

• Tensor-based approaches (Grefenstette et al., 2011; Van de Cruys et al., 2013; Milajevs et al., 2014)
  – Effective in transitive verb disambiguation
  – Composition functions
    • Not learned, but computed in postprocessing

• Joint learning approach (Hashimoto et al., 2014)
  – Word embeddings and composition functions
    • Jointly learned from scratch (w/o word2vec!)
  – Interaction between verbs and their arguments
    • Very weak
An Implicit Tensor Factorization Method

- Bridging the gap between tensor-based and joint learning approaches

Implicit factorization method (Levy and Goldberg, 2014)

Implicit tensor factorization (this work)

State-of-the-art result on a verb sense disambiguation task!
Today’s Agenda

1. Introduction

2. Related Work
   – Joint learning and tensor-based approaches

3. Learning Embeddings for Transitive Verb Phrases
   – The Role of Prepositional Adjuncts
   – Implicit Tensor Factorization

4. Experiments and Results

5. Summary
Approaches to Phrase Embeddings

- **Element-wise addition/multiplication** (Mitchell and Lapata, 2010)
  - $v(\text{sentence}) = \sum_i v(w_i)$

- **Recursive autoencoders**
  - Using parse trees (Socher et al., 2011; Hermann and Blunsom, 2013)
    - $v(\text{parent}) = f(v(\text{left child}), v(\text{right child}))$

- **Tensor/matrix-based methods**
  - $v(\text{adj noun}) = M(\text{adj})v(\text{noun})$ (Baroni and Zamparelli, 2010)
  - $M(\text{verb}) = \sum_{i,j} v(\text{subj}_i)^T v(\text{obj}_j)$ (Grefenstette and Sadrzadeh, 2011)
    - $M(\text{subj}, \text{verb}, \text{obj}) = \{v(\text{subj})^T v(\text{obj})\} \times M(\text{verb})$
    - $v(\text{subj}, \text{verb}, \text{obj}) = \{M(\text{verb})v(\text{obj})\} \times v(\text{subj})$ (Kartsaklis et al., 2012)
Which Word Embeddings are the Best?

- Co-occurrence matrix + SVD, NMF, etc.
- C&W (Collobert and Weston, 2011)
- RNNLM (Mikolov et al., 2013)
- SkipGram/CБOW (Mikolov et al., 2013)
- vLBL/ivLBL (Mnih and Kavukcuoglu, 2013)
- Dependency-based SkipGram (Levy and Goldberg, 2014)
- Glove (Pennington et al., 2014)

Which word embeddings should we use for which composition methods?

Joint leaning
Co-Occurrence Statistics of Phrases

- Word co-occurrence statistics $\rightarrow$ word embeddings
- How about phrase embeddings?
  - Phrase co-occurrence statistics!

The importer **made payment** in his own domestic currency

The businessman **pays his monthly fee** in yen

Similar contexts

Similar meanings?
Today’s Agenda

1. Introduction
2. Related Work
   – Joint learning and tensor-based approaches
3. Learning Embeddings for Transitive Verb Phrases
   – The Role of Prepositional Adjuncts
   – Implicit Tensor Factorization
4. Experiments and Results
5. Summary
How to Identify Phrase-Word Relations?

- Using predicate-argument structures (Hashimoto et al., 2014)
  - *Enju* parser (Miyao et al., 2008)
  - Analyzes relations between phrases and words

```
The importer made payment in his own domestic currency
```

```markdown
Arguments

NP

The importer
made
payment

Verb

Prepositions

Prepositions

Adjunct
```
Training Data from Large Corpora

• Focusing on the role of **prepositional adjuncts**
  – Prepositional adjuncts **complement meanings** of verb phrases → should be useful

How to model the relationships between predicates and arguments?

---

Parse

English Wikipedia, BNC, etc.

Simplification

31/07/2015 CVSC2015 in Beijing, China
Today’s Agenda

1. Introduction
2. Related Work
   - Joint learning and tensor-based approaches
3. Learning Embeddings for Transitive Verb Phrases
   - The Role of Prepositional Adjuncts
   - Implicit Tensor Factorization
4. Experiments and Results
5. Summary
Tensor-Based Approaches

- Tensor/matrix-based approaches (Noun: vector)
  - Transitive verb: matrix

\[ \text{PMI}((\text{importer, make, payment}) = 0.31 \]

(Grefenstette and Sadrzadeh, 2011; Van de Cruys et al., 2013)
Implicit Tensor Factorization (1)

- Parameterizing
  - **Predicate matrices** and argument embeddings
  - Similar to an implicit matrix factorization method for learning word embeddings (Levy and Goldberg, 2014)
Implicit Tensor Factorization (2)

- Calculating plausibility scores
  - Using predicate matrices & argument embeddings

\[ T(p, a_1, a_2) = \]
Implicit Tensor Factorization (3)

- Learning model parameters
  - Using plausibility judgment task
- Observed tuple: \((p, a_1, a_2)\)
- Collapsed tuples: \((p', a_1, a_2), (p, a_1', a_2), (p, a_1, a_2')\)
  - Negative sampling (Mikolov et al., 2013)

**Cost function**

\[
\begin{align*}
\text{Larger} & \quad - \log \sigma(T(p, a_1, a_2)) - \log(1 - \sigma(T(p', a_1, a_2))) \\
\text{Smaller} & \quad - \log(1 - \sigma(T(p, a_1', a_2))) - \log(1 - \sigma(T(p, a_1, a_2'))) 
\end{align*}
\]
Example

- Discriminating between observed and collapsed ones

\[
(p, a_1, a_2) = (\text{in}, \text{importer make payment}, \text{currency})
\]
\[
(p', a_1, a_2) = (\text{on}, \text{importer make payment}, \text{currency})
\]
\[
(p, a_1', a_2) = (\text{in}, \text{child eat pizza}, \text{currency})
\]
\[
(p, a_1, a_2') = (\text{in}, \text{importer make payment}, \text{furniture})
\]
How to Compute SVO Embeddings?

- Two methods:
  - (a) assigning a vector to each SVO tuple
  - (b) composing SVO embeddings

- Parameterized matrices
- Parameterized vectors
- Composed vectors

[Kartsaklis et al., 2012]
Today’s Agenda

1. Introduction
2. Related Work
   – Joint learning and tensor-based approaches
3. Learning Embeddings for Transitive Verb Phrases
   – The Role of Prepositional Adjuncts
   – Implicit Tensor Factorization
4. Experiments and Results
5. Summary
Experimental Settings

- Training corpus (English Wikipedia)
  - SVO data: 23.6 million instances
  - SVO-preposition-noun data: 17.3 million instances
- Parameter initialization
  - Random values
- Optimization
  - Mini-batch *AdaGrad* (Duchi et al., 2011)
- Embedding dimensionality
  - 50

*How do we tune the parameters?*
*For more details, please come to see the poster session!*
Examples of Learned SVO Embeddings

- Composing SVO embeddings

|                      | Nearest neighbor verb-object phrases                                      |
|----------------------|---------------------------------------------------------------------------|
| make money           | make cash, make dollar, make profit, earn baht, earn pound, earn billion   |
| make payment         | make loan, make repayment, pay fine, pay amount, pay surcharge, pay reimbursement |
| make use (of)        | use number, use concept, use approach, use method, use model, use one     |

Capturing the changes of the meaning of “make”
The learned verb matrices capture multiple meanings:

| verb  | nearest neighbors                                      |
|-------|--------------------------------------------------------|
| run   |                                                        |
| 27th  | operate, execute, insert, hold, grid, produce, add, assume, manage, render |
| col.  |                                                        |
| 34th  | release, operate, create, override, govern, oversee, distribute, host, organize |
| row   |                                                        |
| all   | operate, start, manage, own, launch, continue, establish, open, maintain |
| encode|                                                        |
| 28th  | denature, transfect, phosphorylate, polymerize, subtend, acid |
| row   |                                                        |
| 39th  | format, store, decode, embed, concatenate, encrypt, memorize |
| row   |                                                        |
| all   | concatenate, permute, phosphorylate, quantize, composite, transfect, transduce |
Verb Sense Disambiguation Task

- Measuring semantic similarities of verb pairs taking the same subjects and objects \((\text{Grefenstette and Sadrzadeh, 2011})\)
  - Evaluation: Speaman’s rank correlation between similarity scores and human ratings

| Verb pair with subj&obj | Human rating |
|-------------------------|-------------|
| student **write** name  | 7           |
| student **spell** name  |             |
| child **show** sign     | 6           |
| child **express** sign  |             |
| system **meet** criterion |           |
| system **visit** criterion |         |
Results

- State-of-the-art results on the disambiguation task
  - Prepositional adjuncts improve the results

| Method                                      | Spearman’s rank correlation score |
|---------------------------------------------|-----------------------------------|
| This work (only verb data)                  | 0.480                             |
| This work (verb and preposition data)       | 0.614                             |
| Tensor-based approach (Milajevs et al., 2014)| 0.456                             |
| Joint learning approach (Hashimoto et al., 2014)| 0.422                             |

For more details, please come to see the poster session!
Today’s Agenda

1. Introduction
2. Related Work
   – Joint learning and tensor-based approaches
3. Learning Embeddings for Transitive Verb Phrases
   – The Role of Prepositional Adjuncts
   – Implicit Tensor Factorization
4. Experiments and Results
5. Summary
Summary

• Word and phrase embeddings are jointly learned using large corpora parsed by syntactic parsers
  – Tensor-based method is suitable for verb sense disambiguation
  – Adjuncts are useful in learning verb phrases

• Future directions:
  – improving the embedding methods
  – applying them to real-world NLP applications
    • What kind of information should be captured?