Learning Embeddings for Transitive Verb Disambiguation by Implicit Tensor Factorization

Kazuma Hashimoto

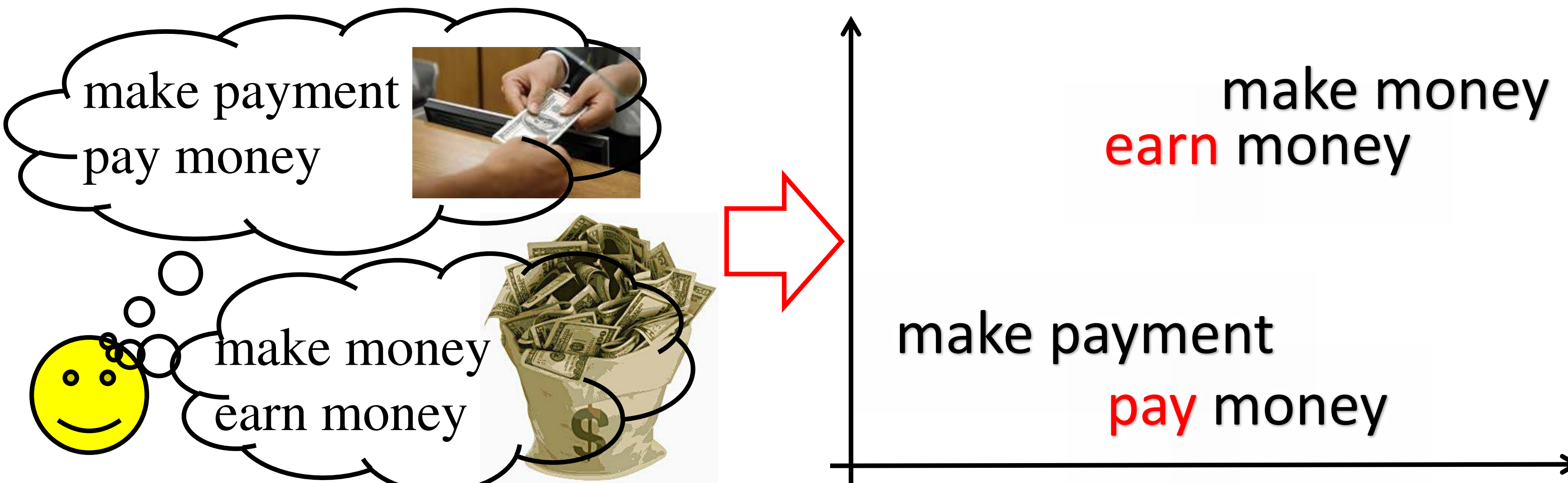
Yoshimasa Tsuruoka

University of Tokyo

31/07/2015 CVSC2015 in Beijing, China

Composition: Words → Phrases

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



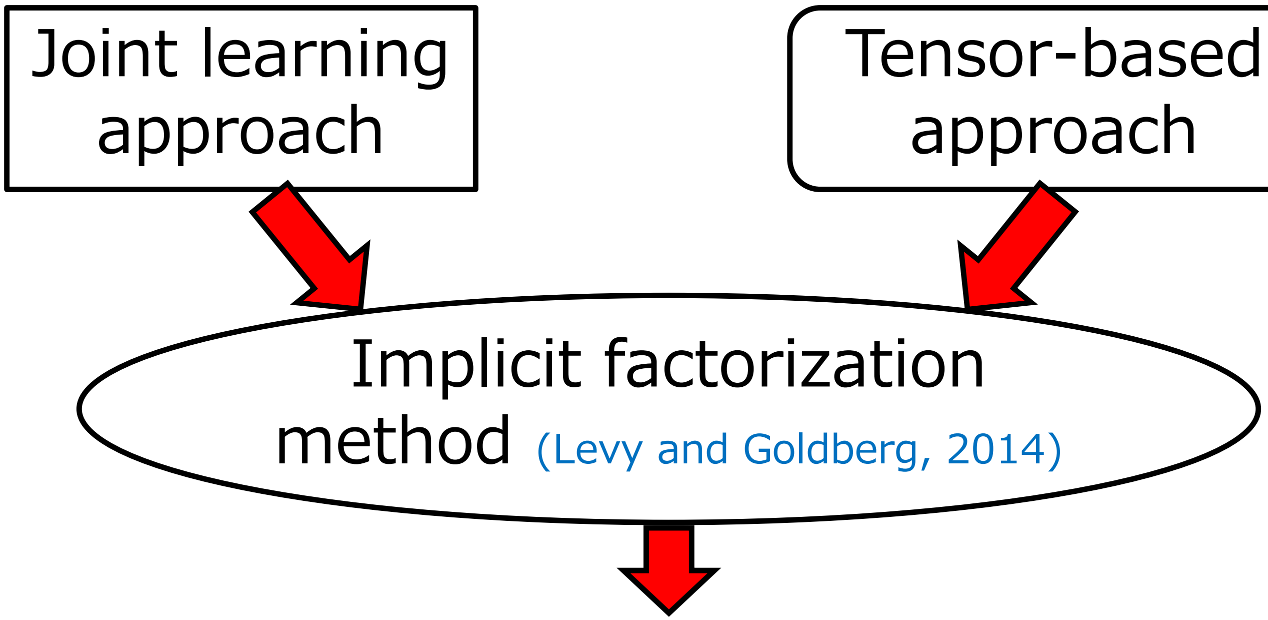
Embeddings of Transitive Verb Phrases

- 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 tensor factorization (this work)

<u>State-of-the-art result</u>
<u>on a verb sense disambiguation task!</u>

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)
  - $v(\text{sentnce}) = \sum_i v(w_i)$
- 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)
  - $M(\text{verb}) = \sum_{i,j} v(subj_i)^{\mathrm{T}} v\left(obj_j\right)$ (Grefenstette and Sadrzadeh, 2011)
    - $M(\text{subj, verb, obj})=\{v(\text{subj})^{\mathrm{T}} v(\text{obj})\} * M(\text{verb})$
    - $v(\text{subj, verb, obj})=\{M(\text{verb})v(\text{obj})\} * v(\text{subj})$ (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/CBOW (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?

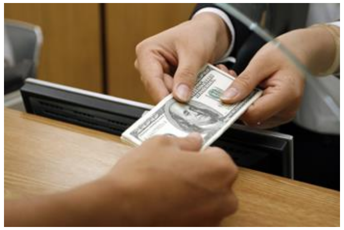
⇨ ***Joint*** **leaning**

Co-Occurrence Statistics of Phrases

- Word co-occurrence statistics → word embeddings
- How about phrase embeddings?
  - **Phrase co-occurrence statistics!**

Similar meanings?

The businessman **<u>pays his monthly fee</u>** in yen



Similar contexts

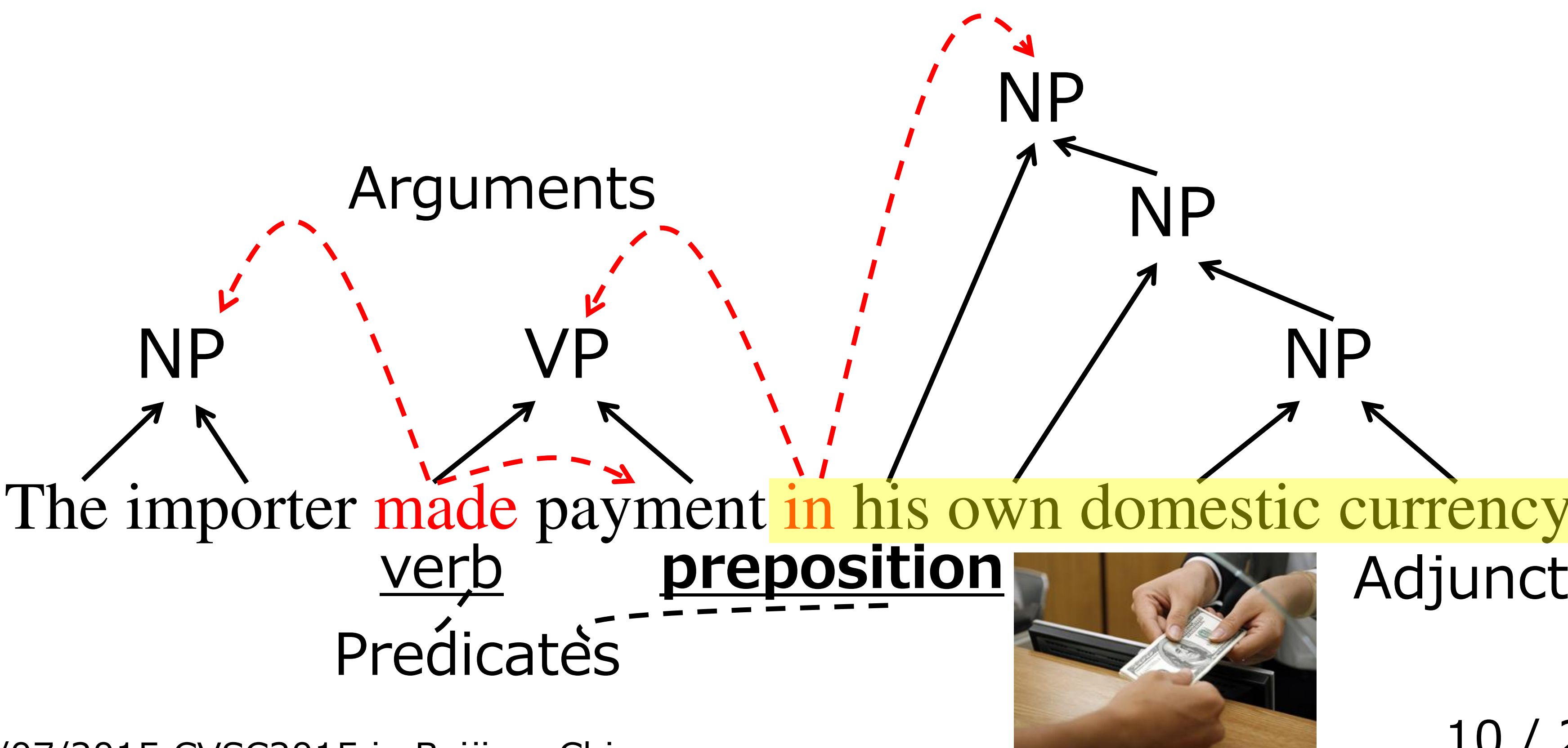
The importer **<u>made payment</u>** in his own domestic currency

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?

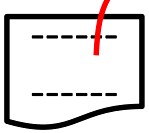
- Using predicate-argument structures (Hashimoto et al., 2014)
  - *Enju* parser (Miyao et al., 2008)
    - Analyzes relations between phrases and words



Training Data from Large Corpora

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

Parse



English Wikipedia, BNC, etc.

predicate	argument 1	argument 2
make	an importer	payment
in	make payment	his own domestic currency

Simplification

predicate	argument 1	argument 2
make	importer	payment
in	importer make payment	currency

How to model the relationships between predicates and arguments?

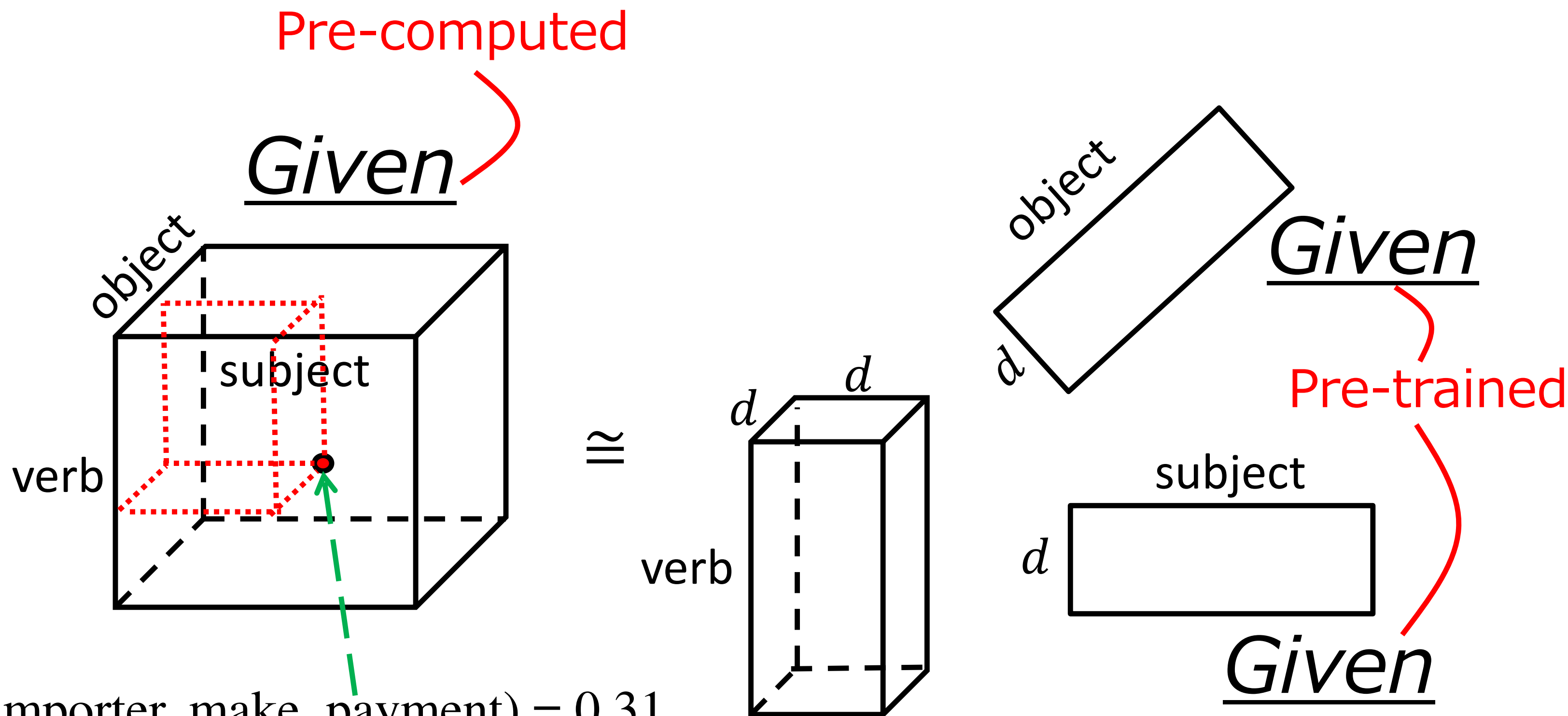
Today's Agenda

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**

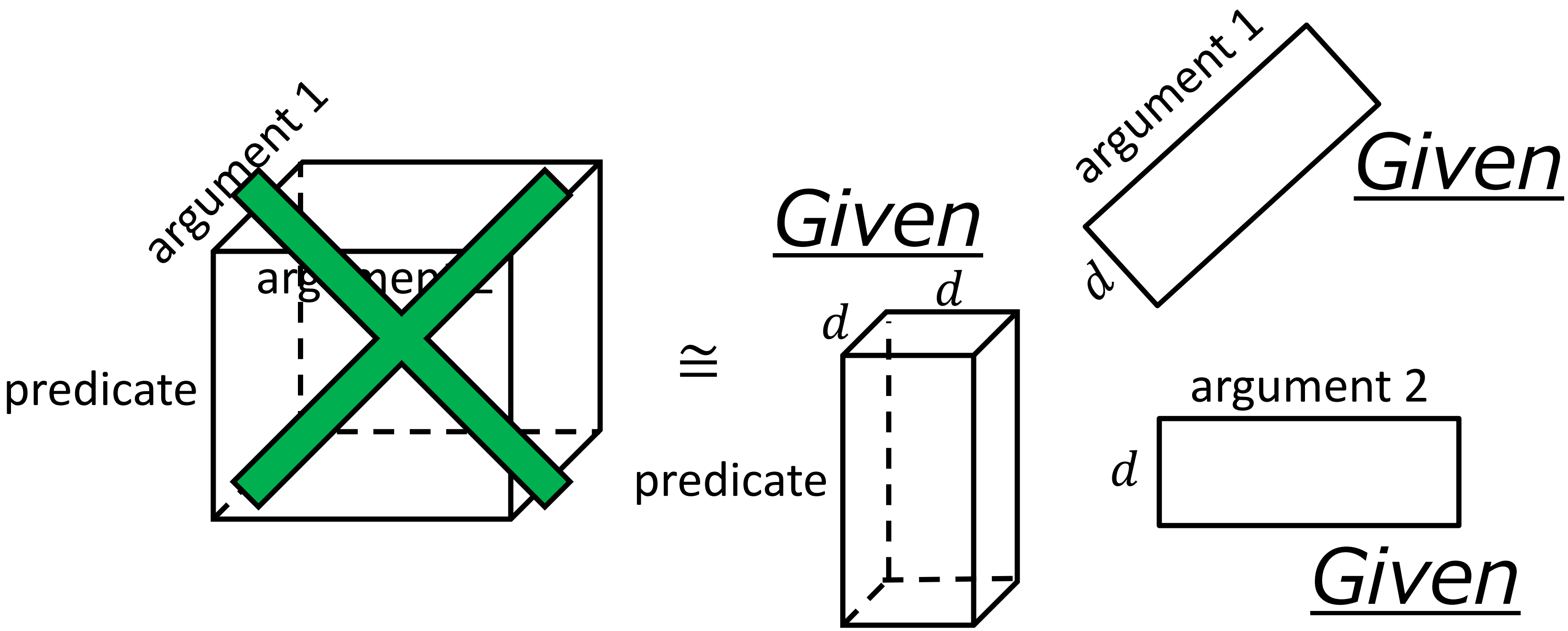
(Grefenstette and Sadrzadeh, 2011; Van de Cruys et al., 2013)



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

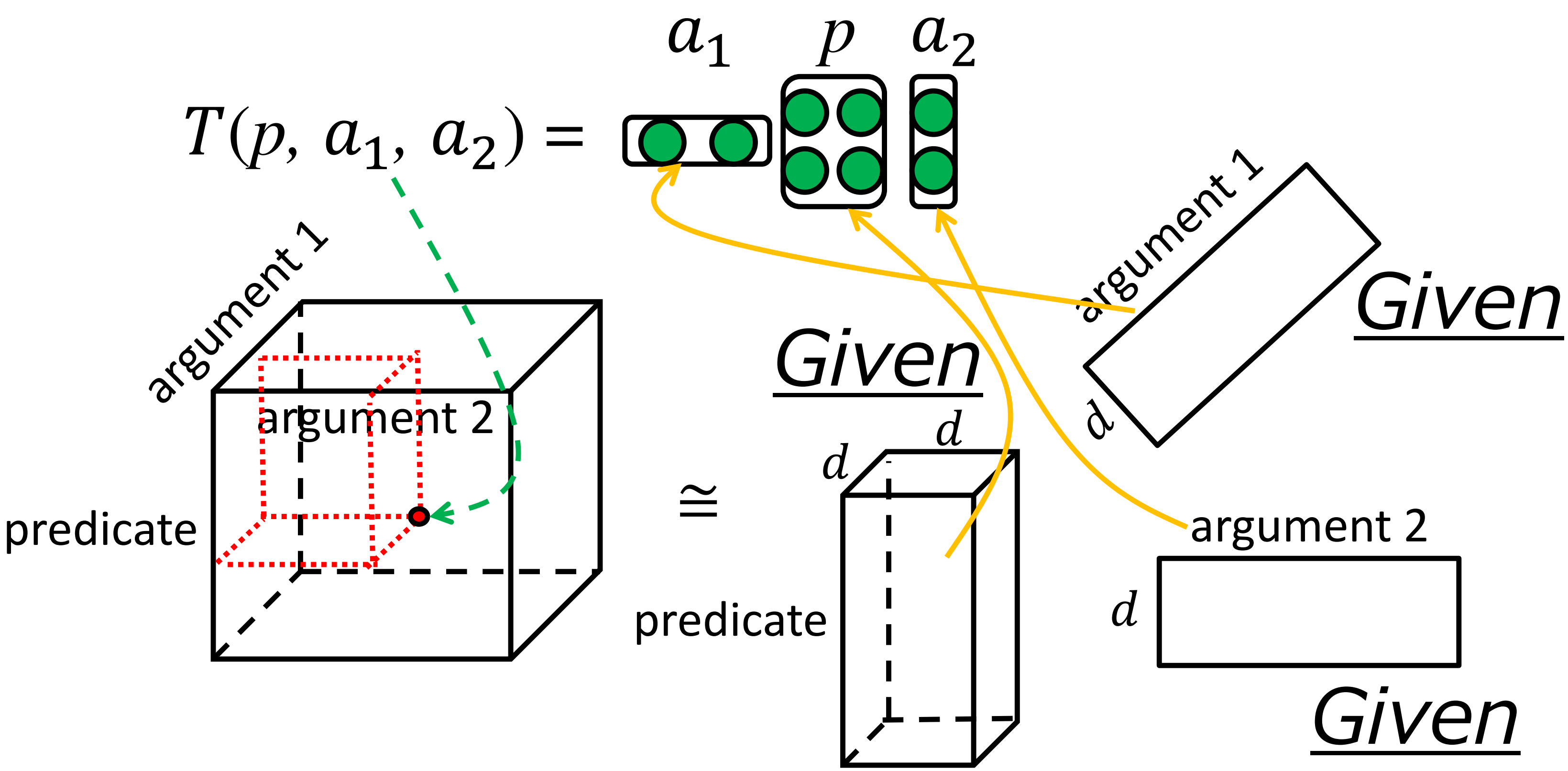
Implicit Tensor Factorization (1)

- 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



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

$$-\log \sigma(\underbrace{T(p, a_1, a_2)}_{\text{Larger}}) - \log\left(1 - \sigma(\underbrace{T(p', a_1, a_2)}_{\text{Smaller}})\right)$$
$$- \log\left(1 - \sigma(T(p, a_1', a_2))\right)$$
$$- \log\left(1 - \sigma(T(p, a_1, a_2'))\right)$$

Example

- Discriminating between observed and collapsed ones

predicate	argument 1	argument 2
make	importer	payment
in	importer make payment	currency

$(p,\ a_1,\ a_2)$ = (in, importer make payment, currency)

$(p',\ a_1,\ a_2)$= (*on*, importer make payment, currency)

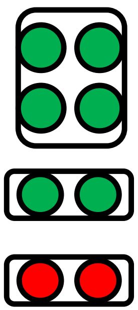
$(p,\ a_1',\ a_2)$= (in, *child eat pizza*, currency)

$(p,\ a_1,\ a_2')$= (in, importer make payment, *furniture*)

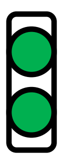
$$-\log\sigma\Big(\underbrace{T(p,\ a_1,\ a_2)}_{\text{Larger}}\Big) - \log\Big(1-\sigma\big(\underbrace{T(p',\ a_1,\ a_2)}_{\text{Smaller}}\big)\Big)$$
$$-\log\Big(1-\sigma\big(T(p,\ a_1',\ a_2)\big)\Big)$$
$$-\log\Big(1-\sigma\big(T(p,\ a_1,\ a_2')\big)\Big)$$

How to Compute SVO Embeddings?

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

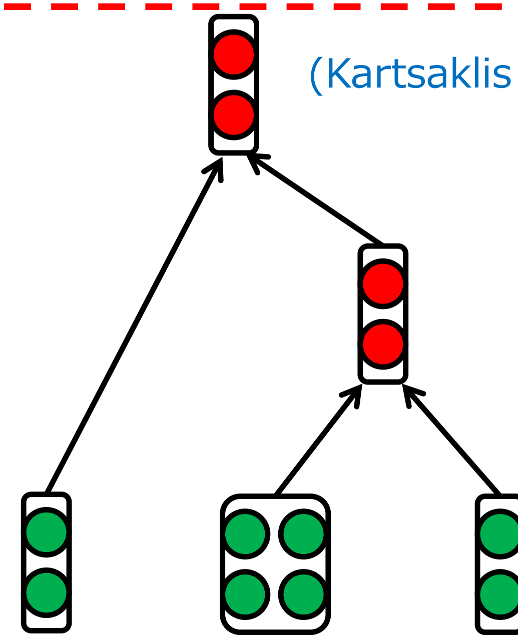


- Parameterized matrices
- Parameterized vectors
- Composed vectors



[importer make payment]

(a)



(Kartsaklis et al., 2012)

[importer make payment]

(b)

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”

Multiple Meanings in Verb Matrices

- The learned verb matrices capture multiple meanings

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

Different usage

Mixed (Similar to word embeddings)

Verb Sense Disambiguation Task

- Measuring semantic similarities of verb pairs taking the same subjects and objects (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 student **spell** name	7
child **show** sign child **express** sign	6
system **meet** criterion system **visit** criterion	1

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?