Translation Similarity Model Based on Bilingual Compositional Semantics

Wang Chaochao，Xiong Deyi，Zhang Min

Soochow University

December 8, 2014

Outline

1. Background
2. Motivation
3. Translation Similarity Model
4. Experiments
5. Conclusion and Future Work

Background

1. Distributional Hypothesis
   Occurring in similar contexts,having same meaning (Harris,1968)
2. Distributional Semantics
3. Challenging and Important.
4. Helpful to phrase-based SMT

Example

Woman	100	200	0	500	50
Vocabulary	money	cosmetic	beard	shopping	drinking
Man	150	10	200	50	150

Compositional Semantics

1. Principle
   - meanings
   - rules
2. Methods
   - λ -Calculus
   - **Vector Addition**
   - other deep learning methods

Why does SMT need Compositional Semantics

Example

- Chinese Sentence: 小张说故事很有趣.

Different Composition,Different Meanings.

Why does SMT need Compositional Semantics

Example

- Chinese Sentence: 小张说故事很有趣.
- Composition 1: 小张 说 故事 很 有趣.
- Translation 1: Zhang said the story is very interesting .

Different Composition,Different Meanings.

Why does SMT need Compositional Semantics

Example

- Chinese Sentence: 小张说故事很有趣.
- Composition 1: 小张 说 故事 很 有趣.
- Translation 1: Zhang said the story is very interesting .
- Composition 2: 小张 说故事 很 有趣.
- Translation 2: Zhang telling story is very interesting.

Different Composition,Different Meanings.

Motivation

Pevious Approaches

1. Word-level Semantics
   - Mikolov et al. (2013) : **Transformation Matrix**
   - Zou et al. (2013) : **Bilingual Training**
2. Phrase-level Semantics
   - Gao et al. (2014) : **Jointly Traing**

Problems

1. Still focus on **linear projection**

Motivation

Pevious Approaches

1. Word-level Semantics
   - Mikolov et al. (2013) : **Transformation Matrix**
   - Zou et al. (2013) : **Bilingual Training**
2. Phrase-level Semantics
   - Gao et al. (2014) : **Jointly Traing**

Problems

1. Still focus on **linear projection**
2. Defined at **word level**

Motivation

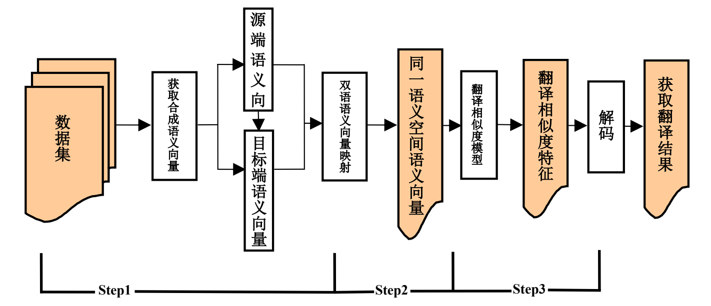
Pevious Approaches

1. Word-level Semantics
   - Mikolov et al. (2013) : **Transformation Matrix**
   - Zou et al. (2013) : **Bilingual Training**
2. Phrase-level Semantics
   - Gao et al. (2014) : **Jointly Traing**

Problems

1. Still focus on **linear projection**
2. Defined at **word level**
3. **Jointly learning** limits application in SMT

Framework



Step1: Monolingual Compositional Semantics

Step2: Projection

Step3: Calculating and Integrating

Translation Similarity Model

Goal :

Learning Compositional Representations

How?

1. Point-Wise Mutual Information based vector space word representations

$$pmi(c,t) = log \frac{p(c,t)}{p(c)p(t)}$$

2. Weighted Vector Addition

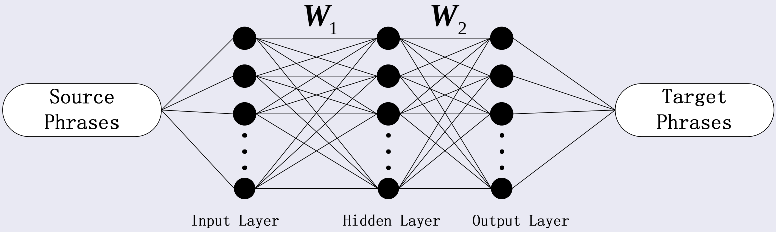
$$\vec{p} = \alpha \vec{w_1} + \beta \vec{w_2}$$

Translation Similarity Model

Goal :

Learning Bilingual Representations

Project



Goal :

Calculating Similarity

How?

1. Between Two Phrases

$$Sim(p(\overrightarrow{w}_1), \vec{w}_2) = \frac{p(\overrightarrow{w}_1) \bullet \vec{w}_2}{\left\| p(\overrightarrow{w}_1) \right\| \times \| \vec{w}_2 \|}$$

2. Source Sentence

$$M_{sim} = \prod_{(w_1, w_2) \in P} sim(p(\overrightarrow{w}_1), \overrightarrow{w}_2)$$

How to Integrate into SMT?

As a feature in any SMT systems.

Log Linear Model

1. General Log Linear Model

$$Pr(e|f) \approx P_{\lambda_1 \dots \lambda_M}(e|f) = \frac{exp[\sum \lambda_m h_m(e,f)]}{\sum exp[\lambda_m h_m(e',f))]}$$

2. New Feature M_{sim}

$$M_{sim} = \prod_{(w_1,w_2)\in P} sim(p(\overrightarrow{w}_1), \overrightarrow{w}_2)$$

Experiment Setup

1. Language: Chinese $\rightarrow$ English
2. Baseline system: Hierarchical Phrase-based System
3. Training data: 4.1M sentence pairs from LDC data
4. Development data: NIST 2005
5. Test data: NIST 2006 and NIST 2008
6. Open Source Toolkit: DISSECT
   (To get word representations and weights)

Results

Systems	Development Set	Test Set	
	NIST 05	NIST 06	NIST 08
Baseline	33.53	30.66	23.32
Non-projection	32.76	30.38	23.12
Linear Projection	34.62	30.77	23.56
Non-linear Projection	34.53	**31.22**	**23.74**

Analysis

1. Compositional semantics can improve translation quality

Results

Systems	Development Set	Test Set	
	NIST 05	NIST 06	NIST 08
Baseline	33.53	30.66	23.32
Non-projection	32.76	30.38	23.12
Linear Projection	34.62	30.77	23.56
Non-linear Projection	34.53	**31.22**	**23.74**

Analysis

1. Compositional semantics can improve translation quality
2. It is necessary to learn bilingual phrase representation

Results

Systems	Development Set	Test Set	
	NIST 05	NIST 06	NIST 08
Baseline	33.53	30.66	23.32
Non-projection	32.76	30.38	23.12
Linear Projection	34.62	30.77	23.56
Non-linear Projection	34.53	**31.22**	**23.74**

Analysis

1. Compositional semantics can improve translation quality
2. It is necessary to learn bilingual phrase representation
3. Non-linear projection is more effective than linear projection

Conclusions

1. We presented a flexible framework to learn bilingual distributed compositional representations for SMT

Conclusions

1. We presented a flexible framework to learn bilingual distributed compositional representations for SMT
2. Our experiments validate that Non-linear projection is more powerful than linear projection between source and target language semantic space

Future Work

1. Using deep learning to learn word and compositional representations

Future Work

1. Using deep learning to learn word and compositional representations
2. Finding better projection models

Future Work

1. Using deep learning to learn word and compositional representations
2. Finding better projection models
3. Speeding up decoder

Thank You!