Translating Phrases in Neural Machine Translation

Xing Wang[1] (王 星),
Joint work with Zhaopeng Tu[2], Deyi Xiong[1], Min Zhang[1]
[1]Soochow Univesity & [2]Tencent AI Lab

Outline

- Motivation
- Related Work
- Approach
- Experiments
- Conclusion

Outline

- Motivation
- Related Work
- Approach
- Experiments
- Conclusion

Motivation

- Phrases play an important role in natural language understanding and machine translation.
  - Neural machine translation is an approach to machine translation that uses a large neural network.

- However, the word-by-word generation philosophy in NMT makes it difficult to translate multi-word phrases.

Motivation

- Phrases are much better than words as translation units in SMT and have made a significant advance in translation quality.

- Can we translate phrases in NMT?

Outline

- Motivation
- Related Work
- Approach
- Experiments
- Conclusion

Related Work

- the Combination of NMT and SMT

1st Workshop on Neural Machine Translation (and Generation)

Search this site

OVERVIEW
ACCEPTED PAPERS
CALL FOR PAPERS
ORGANIZATION
SPEAKERS
SITEMAP

Call for Papers

The workshop is broad in scope and invites original research contributions on topics that include, but are not limited to the following:

- Incorporating linguistic insights: syntax, alignment, reordering, etc.
- Combining NMT & SMT
- Handling resource-limited domains
- Utilizing more data in NMT: monolingual, multilingual resources
- Multi-task learning for NMT
- NMT for mobile devices
- Analysis and visualization of NMT models
- Beyond sentence-level translation
- Beyond maximum-likelihood estimation
- **Neural Machine Generation**

Related Work

- Sentence-Level Combination:
  - Pre-Translation for Neural Machine Translation (COLING 2016)
  - Neural System Combination for Machine Translation (ACL2017)

Related Work

- Word-Level Combination:
  - Improved Neural Machine Translation with SMT Features (AAAI2016)
  - Syntactically Guided Neural Machine Translation (ACL2016)
  - Incorporating Discrete Translation Lexicons into Neural Machine Translation (EMNLP2016)
  - Neural Machine Translation Advised by Statistical Machine Translation (AAAI2017)

Related Work

- Phrase-Level Combination:

Related Work

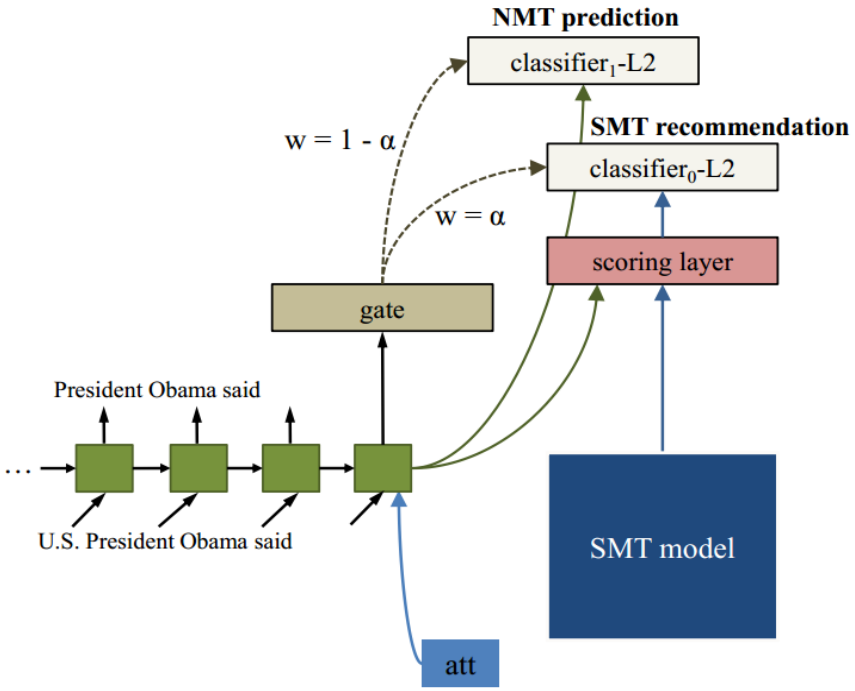
- Phrase-Level Combination:
  - Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search (EMNLP2017)

Outline

- Motivation
- Related Work
- Approach
- Experiments
- Conclusion

Approach

- Previous work: Neural Machine Translation Advised by Statistical Machine Translation(AAAI2017)



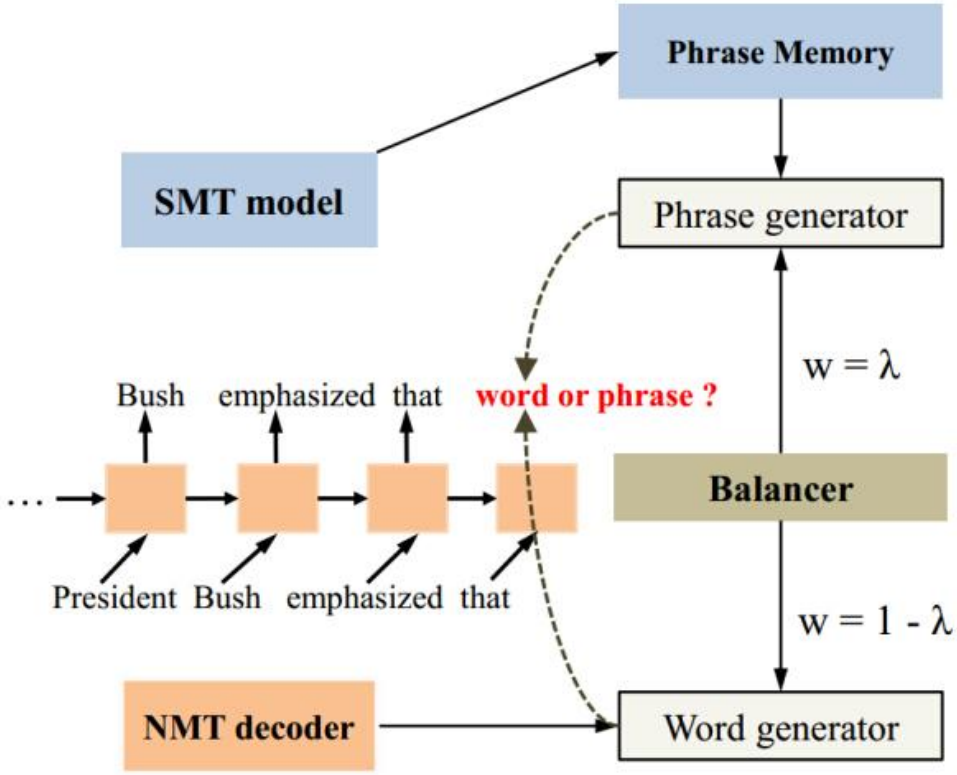
Approach

- Previous work: Neural Machine Translation Advised by Statistical Machine Translation(AAAI2017)

  - SMT word Recommendations
  - Scoring Layer
  - Gating Mechanism/Combination

Approach

- Proposed Architecture



Approach

- Phrase Memory

  ➢Writes to Phrase Memory

  SMT model writes *relevant target phrases* to the phrase memory.

  ➢Reads Phrase Memory

  NMT model reads and scores the target phrases in the phrase memory.

Approach

- Relevant target phrases
- Source: 布什总统强调美国政府坚持一个中国政策、遵守中美三个联合公报。
- NMT partial translation: *President Bush emphasized that* □
- Relevant phrases (Adequacy & Coverage):

  √

  ×

Approach

- Relevant target phrases
- Source: 布什总统强调美国政府坚持一个中国政策、遵守中美三个联合公报。
- NMT partial translation: *President Bush emphasized that* □
- Relevant phrases (Adequacy & Coverage):

  √ the United States

  ×

Approach

- Relevant target phrases
- Source: 布什总统强调美国政府坚持一个中国政策、遵守中美三个联合公报。
- NMT partial translation: *President Bush emphasized that* □
- Relevant phrases (Adequacy & Coverage):

  √ the United States, the One-China policy

  ×

Approach

- Relevant target phrases
- Source: 布什总统强调美国政府坚持一个中国政策、遵守中美三个联合公报。
- NMT partial translation: *President Bush emphasized that* □
- Relevant phrases (Adequacy & Coverage):

  √ the United States, the One-China policy

  × President Bush, The Iraq War

Approach

- Source: [布什总统] 强调 [美国政府] 坚持 [一个中国政策]、遵守[中美三个联合公报]。
- NMT partial translation: *President Bush emphasized that* □

SMT pre-translates the source chunks and scores the target phrases.

Approach

- Source: [布什总统] 强调 [美国政府] 坚持 [一个中国政策]、遵守[中美三个联合公报]。
- NMT partial translation: *President Bush emphasized that* □

SMT pre-translates the source chunks and scores the target phrases.

SMT writes relevant phrases into Phrase Memory：

the United States,
the United States of America
the One-China policy
…….

Approach

- Source: [布什总统] 强调 [美国政府] 坚持 [一个中国政策]、遵守[中美三个联合公报]。
- NMT partial translation: *President Bush emphasized that* □

SMT pre-translates the source chunks and scores the target phrases.

SMT writes relevant phrases into Phrase Memory:

the United States,
the United States of America
the One-China policy
…….

NMT reads Phrase Memory and scores phrases. NMT chooses to generate word or phrases in Phrase Memory.

Approach

- NMT reads Phrase Memory and scores phrases. NMT chooses to generate word or phrases in Phrase Memory.

- Neural network based balancer:
  - The trade-off between word generation mode and phrase generation mode is balanced by a weight λ.

Approach

- The probability of generating y is calculated by

$$P(\mathbf{y}|\mathbf{x}) = \prod_{w_k \in \mathbf{w}} (1 - \lambda_{t(w_k)}) P_{word}(w_k) \times \prod_{p_l \in \mathbf{p}} \lambda_{t(p_l)} P_{phrase}(p_l)$$

- where balancing weight λ is computed by

$$\lambda_i = \sigma(f_b(s_i, y_{i-1}, c_i))$$

Outline

- Motivation
- Related Work
- Approach
- Experiments
- Conclusion

Experiments

- 1.25M LDC data, NIST test sets (Chinese-English)

SYSTEM	NIST04	NIST05	NIST08	Avg
Moses	34.74	31.99	23.69	30.14
RNNSearch	37.80	34.70	24.93	32.48
+memory	38.21	35.15	25.48†	32.95
+memory +chunking tag	38.83‡	35.72‡	26.09‡	33.55

- Systems:
  - Moses (phrase-based SMT model): 4-gram language models using KenLM; GIZA++; minimum error rate.
  - RNNSearch (NMT Baseline): Bahdanau et al., 2015
  - Our model is integrated into RNNSearch

Experiments

- Syntactic Categories of Generated Phrases

Type	All		New	
	Total	Correct	Total	Correct
NP	81.0%	38.7%	46.0%	11.5%
VP	8.0%	1.7%	6.5%	0.8%
QP	10.8%	4.1%	6.2%	0.9%
Others	0.2%	0%	0.2%	0%
Sum	100%	44.5%	58.9%	13.2%

  - Percentages of phrase categories to the total number of generated ones.

Experiments

- Number of Words in Generated Phrases

Words	All		New	
	Total	Correct	Total	Correct
2	66.2%	33.6%	34.9%	9.1%
3	20.7%	8.4%	13.4%	3.2%
4	7.4%	1.9%	5.4%	0.6%
⩾5	5.7%	0.6%	5.2%	0.3%

  - Percentages of phrases with different word counts to the total number of generated ones.

Experiments

- Limitations:
  - SMT translation
  - Chunking error
  - Word generation

Outline

- Motivation
- Related Work
- Approach
- Experiments
- Conclusion

Conclusion

Backbone	Word-level	Phrase-level	?
Shallow			
Deep			

- Improved Neural Machine Translation with SMT Features (AAAI2016)
  - Neural Machine Translation Advised by Statistical Machine Translation (AAAI2017)
    - Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search (EMNLP2017)
      - Translating Phrases in Neural Machine Translation (EMNLP2017)

Conclusion

Backbone	Word-level	Phrase-level	?
Shallow	AAAI2016		
Deep			

- Improved Neural Machine Translation with SMT Features (AAAI2016)
  - Neural Machine Translation Advised by Statistical Machine Translation (AAAI2017)
    - Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search (EMNLP2017)
      - Translating Phrases in Neural Machine Translation (EMNLP2017)

Conclusion

Backbone	Word-level	Phrase-level	?
Shallow	AAAI2016		
Deep	AAAI2017		

- Improved Neural Machine Translation with SMT Features (AAAI2016)
  - Neural Machine Translation Advised by Statistical Machine Translation (AAAI2017)
    - Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search (EMNLP2017)
      - Translating Phrases in Neural Machine Translation (EMNLP2017)

Conclusion

Backbone	Word-level	Phrase-level	?
Shallow	AAAI2016	EMNLP2017	
Deep	AAAI2017		

- Improved Neural Machine Translation with SMT Features (AAAI2016)
- Neural Machine Translation Advised by Statistical Machine Translation (AAAI2017)
- Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search (EMNLP2017)
- Translating Phrases in Neural Machine Translation (EMNLP2017)

Conclusion

Backbone	Word-level	Phrase-level	?
Shallow	AAAI2016	EMNLP2017	
Deep	AAAI2017	EMNLP2017	

- Improved Neural Machine Translation with SMT Features (AAAI2016)
- Neural Machine Translation Advised by Statistical Machine Translation (AAAI2017)
- Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search (EMNLP2017)
- Translating Phrases in Neural Machine Translation (EMNLP2017)

Conclusion

Backbone	Word-level	Phrase-level	?
Shallow	AAAI2016	EMNLP2017	
Deep	AAAI2017	EMNLP2017	

- Improved Neural Machine Translation with SMT Features (AAAI2016)
- Neural Machine Translation Advised by Statistical Machine Translation (AAAI2017)
- Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search (EMNLP2017)
- Translating Phrases in Neural Machine Translation (EMNLP2017)

Conclusion

Backbone	Word-level	Phrase-level	?
Shallow	AAAI2016	EMNLP2017	?
Deep	AAAI2017	EMNLP2017	?

- Improved Neural Machine Translation with SMT Features (AAAI2016)
  - Neural Machine Translation Advised by Statistical Machine Translation (AAAI2017)
    - Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search (EMNLP2017)
      - Translating Phrases in Neural Machine Translation (EMNLP2017)

Conclusion

- We have presented a novel model to translate source phrases and generate target phrase translations in NMT by integrating the phrase memory into the encoder-decoder architecture.

- Experiment results on Chinese-English translation have demonstrated that the proposed model can improve the translation performance.

Outline

- Motivation
- Related Work
- Approach
- Experiments
- Conclusion

Thank you!

xingwsuda@gmail.com