Multi-reference Training with Pseudo-References for Neural Translation and Text Generation

Renjie Zheng\textsuperscript{1} \hspace{1cm} Mingbo Ma\textsuperscript{1,2} \hspace{1cm} Liang Huang\textsuperscript{1,2}

1 Oregon State University \hspace{1cm} 2 Baidu Research

- an image of a grey tiger
- a black tabby cat sitting curled sleeping
- in a red brown fruit bowl that sits on the kitchen dining room near a window
Multi-reference Data in Image Captioning

- MSCOCO Image Captioning Dataset
  - 114,287 images, each with 5 reference captions
- PASCAL-50S
  - 1,000 images, each with 50 references
- Example from MSCOCO
  - 1: A brown and black cat is sleeping in a bowl on a table
  - 2: A grey tiger cat sleeping in a brown bowl on a table
  - 3: An image of a cat sitting inside of a bowl on the kitchen table
  - 4: A cat asleep in a fruit bowl on a dining room table
  - 5: A gray tabby cat is curled in a red bowl that sits on a table near a window
Multi-reference Data in MT

- ~10,000 Chinese (Arabic) sentences, each with 4 reference English translations
- example from NIST02

1. Indonesia reiterated its opposition to foreign military presence
2. Indonesia repeats its opposition against station of foreign troops in Indonesia
3. Indonesia reiterates opposition to garrisoning foreign armies
4. Indonesia reiterates: no foreign troops in country
Can we use multiple references in training?

- multiple references are always used in evaluations
- previous work under-utilizes available multiple references in training
  - *image captioning*: existing literatures simply sample one of them in each epoch
  - *machine translation*: no previous work using extra references during training
Can we use multiple references in training?

• multiple references are always used in evaluations

• previous work under-utilizes available multiple references in training
  
  • *image captioning*: existing literatures simply sample one of them in each epoch
  
  • *machine translation*: no previous work using extra references during training

• Q: can we benefit more from training with multiple references?
Multi-reference Training
Multi-reference Training

- *First:* only uses first reference

<table>
<thead>
<tr>
<th>Ref.</th>
<th>1&lt;sup&gt;st&lt;/sup&gt;</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt;</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt;</th>
<th>4&lt;sup&gt;th&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Example2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Example3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

First
Multi-reference Training

- **First**: only uses first reference
- **Sample One**: samples one reference for each example in each epoch
Multi-reference Training

- **First**: only uses first reference
- **Sample One**: samples one reference for each example in each epoch
- **Uniform**: utilizes all the references of each example

![Visual Representation]

- **Ref.**
  - Example1
    - 1st: 1
    - 2nd: 1
    - 3rd: 1
    - 4th: 1
  - Example2
    - 1st: 2
    - 2nd: 2
    - 3rd: 2
    - 4th: 2
  - Example3
    - 1st: 3
    - 2nd: 3
    - 3rd: 3
    - 4th: 3

- **First**
  - 1st: 1
  - 2nd: 1
  - 3rd: 1
  - 4th: 1

- **Sample One**
  - 1st: 1
  - 2nd: 2
  - 3rd: 3
  - 4th: 4

- **Uniform**
  - 1st: 1
  - 2nd: 2
  - 3rd: 3
  - 4th: 3

Example 1
Example 2
Example 3

First
Sample One
Uniform
Multi-reference Training

- **First**: only uses first reference
- **Sample One**: samples one reference for each example in each epoch
- **Uniform**: utilizes all the references of each example
- **Shuffle**: also utilizes all the references but shuffles all of them

![Images showing First, Sample One, Uniform, and Shuffle methods]

<table>
<thead>
<tr>
<th>Ref.</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Example2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Example3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

**First**: Only uses the first reference for each example.

**Sample One**: Samples one reference for each example in each epoch.

**Uniform**: Utilizes all the references of each example.

**Shuffle**: Utilizes all the references but shuffles them.
Multi-reference Training

- **First**: only uses first reference
- **Sample One**: samples one reference for each example in each epoch
- **Uniform**: utilizes all the references of each example
- **Shuffle**: also utilizes all the references but shuffles all of them
Multi-reference Training Results

Pre-train on 1M sentence pairs
Fine-tune on 4,667 sents w/ 4 refs
Dev results on 616 sents (NIST06)

<table>
<thead>
<tr>
<th># Refs</th>
<th>Method</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-train</td>
<td>fine-tune</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Pre-train</td>
<td>37.4</td>
</tr>
<tr>
<td>1</td>
<td>First</td>
<td>38.6</td>
</tr>
<tr>
<td>1</td>
<td>Sample One</td>
<td>38.8</td>
</tr>
<tr>
<td>1</td>
<td>Uniform</td>
<td>38.8</td>
</tr>
<tr>
<td>1</td>
<td>Shuffle</td>
<td><strong>38.9</strong></td>
</tr>
</tbody>
</table>

Machine Translation

Image Captioning

Train on MSCOCO 113,287 examples
Dev results on 5,000 examples

<table>
<thead>
<tr>
<th># Refs</th>
<th>Method</th>
<th>BLEU</th>
<th>CIDEr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First</td>
<td>26.3</td>
<td>79.1</td>
</tr>
<tr>
<td>5</td>
<td>Sample One</td>
<td>29.0</td>
<td>85.4</td>
</tr>
<tr>
<td>5</td>
<td>Uniform</td>
<td>30.1</td>
<td>89.8</td>
</tr>
<tr>
<td>5</td>
<td>Shuffle</td>
<td><strong>30.4</strong></td>
<td><strong>91.2</strong></td>
</tr>
</tbody>
</table>
But can we go beyond existing references?

- Text generation potentially have exponentially many possible answers.
- 4 or 5 human references are only a tiny fraction of the whole space (Markus & Marcu 2012).
But can we go beyond existing references?

- Text generation potentially have exponentially many possible answers
- 4 or 5 human references are only a tiny fraction of the whole space (Markus & Marcu 2012)
- Q: can we generate more pseudo-references based on the existing references?
- Q: will these extra references improve generation quality?
How to Generate more Pseudo-references?

- first idea: hard alignment
  - step 1: compress the existing references to a lattice by merging identical words
  - step 2: generate pseudo-references by traversing the lattice

- Indonesia reiterated its opposition to foreign military presence
- Indonesia repeats its opposition against station of foreign troops in Indonesia
- Indonesia reiterates opposition to garrisoning foreign armies

Barzilay & Lee 2002;2003
Step 1: Compress References to Lattice

1. initialize

2.1 pick two sents  
2.2 find identical words  
2.3 merge
Step 1: Compress References to Lattice

1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge

(a) Indonesia reiterated its opposition to foreign military presence.

Indonesia repeats its opposition against station of foreign troops in Indonesia.

Indonesia reiterates opposition to garrisoning foreign armies.
Step 1: Compress References to Lattice

1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge
Step 1: Compress References to Lattice

1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge

Indonesia reiterated its opposition to foreign military presence.

Indonesia repeats its opposition against station of foreign troops in Indonesia.

Indonesia reiterates opposition to garrisoning foreign armies.
Step 1: Compress References to Lattice

1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge
Step 1: Compress References to Lattice

1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge
Step 1: Compress References to Lattice

1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge
Step 1: Compress References to Lattice

1. initialize

2.1 pick two sents

2.2 find identical words

2.3 merge
Step 2: Generate Pseudo-references from Lattice

- first Idea: hard alignment
- step 1: compress the references to lattice by merging the identical words
- step 2: generate pseudo-references by traversing the lattice
Step 2: Generate Pseudo-references from Lattice

- **first Idea:** hard alignment
- **step 1:** compress the references to lattice by merging the identical words
- **step 2:** generate pseudo-references by traversing the lattice
  - 33 pseudo-references can be generated in total
Step 2: Generate Pseudo-references from Lattice

- **first Idea:** hard alignment
  - **step 1:** compress the references to lattice by merging the identical words
  - **step 2:** generate pseudo-references by traversing the lattice
    - 33 pseudo-references can be generated in total
    - Indonesia repeats its opposition to foreign military presence
Step 2: Generate Pseudo-references from Lattice

- first Idea: hard alignment
- step 1: compress the references to lattice by merging the identical words
- step 2: generate pseudo-references by traversing the lattice
  - 33 pseudo-references can be generated in total
  - Indonesia repeats its opposition to foreign military presence
  - Indonesia reiterated its opposition to garrisoning foreign armies
Step 2: Generate Pseudo-references from Lattice

- **first Idea: hard alignment**
- **step 1**: compress the references to lattice by merging the identical words
- **step 2**: generate pseudo-references by traversing the lattice
  - 33 pseudo-references can be generated in total
    - Indonesia repeats its opposition to foreign military presence
    - Indonesia reiterated its opposition to garrisoning foreign armies
    - Indonesia reiterates opposition to foreign troops in Indonesia
    - …
can not merge different words with similar meanings
Limitation 2 of Hard Alignment: Polysemy?

- may merge identical words with different meanings in different contexts

wrong pseudo-reference:

Two elephants try to a small entry
How to Solve these Limitations

- **solution**: identify synonyms and polysemes
- Need a semantic similarity matrix

Indonesia
- reiterated
- its
- opposition
- to
- foreign
- military
- presence

Two elephants
- try
- to
- fit
- through
- a
- small
- entry

Inf.

Prep.
Solution: Identify Synonyms and Polysemes

- train bi-directional Language Model
- predict each word by given the surrounding words
- word pair similarity with cosine similarity

Semantic Substitution Matrix

Indonesia
reiterated
its
opposition
to
foreign
military
presence

Indonesia
repeats
its
opposition
against
station
of
foreign
troops
in
Indonesia
Solution: Identify Synonyms and Polysemes

- train bi-directional Language Model
- predict each word by given the surrounding words
- word pair similarity with cosine similarity
- **reiterated** and **repeats** have high similarity score

Semantic Substitution Matrix
Solution: Identify Synonyms and Polysemes

- train bi-directional Language Model
- predict each word by given the surrounding words
- word pair similarity with cosine similarity
- reiterated and repeats have high similarity score
- military and troops have high similarity score

**Semantic Substitution Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Indonesia</th>
<th>repeats</th>
<th>its</th>
<th>opposition</th>
<th>against</th>
<th>station</th>
<th>of</th>
<th>foreign</th>
<th>troops</th>
<th>in</th>
<th>Indonesia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reiterated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>its</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>opposition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>foreign</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>military</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>presence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Solution: Identify Synonyms and Polysemes

- train bi-directional Language Model
- predict each word by given the surrounding words
- word pair similarity with cosine similarity
- reiterated and repeats have high similarity score
- military and troops have high similarity score
- two to’s have low similarity score
Indonesia reiterated its opposition against station of foreign troops in Indonesia.
Iterative Pairwise Soft Alignment

1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge
1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge
1. initialize

2.1 pick two sents
2.2 find aligned words
2.3 merge

Iterative Pairwise Soft Alignment
Iterative Pairwise Soft Alignment

1. initialize
2.1 pick two sents
2.2 find aligned words
2.3 merge
1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge
Iterative Pairwise Soft Alignment

1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge
Iterative Pairwise Soft Alignment

1. initialize

2.1 pick two sents

2.2 find aligned words

2.3 merge
1. initialize

2.1 pick two sents  2.2 find aligned words  2.3 merge

Iterative Pairwise Soft Alignment
213 pseudo.refs generated (180 more than by hard alignment)
Iterative Pairwise Soft Word Alignment

- 213 pseudo-refs generated (180 more than by hard alignment)
Iterative Pairwise Soft Word Alignment

Soft:

- \text{Indonesia} \xrightarrow{\text{reiterated}} \text{repeats} \xrightarrow{\epsilon} \text{its} \xrightarrow{\text{opposition}} \xrightarrow{\text{to}} \text{garrisoning} \xrightarrow{\epsilon} \text{foreign} \xrightarrow{\text{troops}} \xrightarrow{\text{in}} \text{Indonesia} \xrightarrow{\text{military presence}}

- 213 pseudo-refs generated (180 more than by hard alignment)

- Indonesia reiterates opposition to foreign troops

Hard:

- \text{Indonesia} \xrightarrow{\text{reiterated}} \text{repeats} \xrightarrow{\epsilon} \text{its} \xrightarrow{\text{opposition}} \xrightarrow{\text{to}} \text{garrisoning} \xrightarrow{\epsilon} \text{foreign} \xrightarrow{\text{troops}} \xrightarrow{\text{in}} \text{Indonesia} \xrightarrow{\text{military presence}}

- Indonesia reiterates opposition to foreign troops
• 213 pseudo-refs generated (180 more than by hard alignment)
• Indonesia reiterates opposition to foreign troops
• Indonesia repeats opposition to garrisoning foreign armies
Iterative Pairwise Soft Word Alignment

- 213 pseudo-refs generated (180 more than by hard alignment)
- Indonesia reiterates opposition to foreign troops
- Indonesia repeats opposition to garrisoning foreign armies
- ...
213 pseudo-refs generated (180 more than by hard alignment)

Indonesia reiterates opposition to foreign troops

Indonesia repeats opposition to garrisoning foreign armies

keep best $k$ pseudo-refs according to BLEU score over original refs
• original references
  • A grey tabby cat is curled in a red bowl that sits on a table near a window
  • A brown and black cat is sleeping in a bowl on a table
  • A grey tiger cat sleeping in a brown bowl on a table
  • An image of a cat sitting inside of a bowl on the kitchen table
  • A cat asleep in a fruit bowl on a dining room table

<table>
<thead>
<tr>
<th>Rank</th>
<th>Pseudo-reference</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A grey tiger cat sleeping in a brown bowl on a table near a window</td>
<td>100%</td>
</tr>
<tr>
<td>. .</td>
<td>. . .</td>
<td>. . .</td>
</tr>
<tr>
<td>48</td>
<td>A grey tiger cat sleeping in a fruit bowl on a table</td>
<td>97%</td>
</tr>
<tr>
<td>. .</td>
<td>. . .</td>
<td>. . .</td>
</tr>
<tr>
<td>73,274</td>
<td>A grey and tabby cat inside of a red bowl on the dining room table</td>
<td>0%</td>
</tr>
</tbody>
</table>
Analysis of Generated References

NIST Machine Translation

MSCOCO Image Captioning
Results on Validation Sets

NIST Machine Translation

MSCOCO Image Captioning
Results on Test set

NIST Machine Translation

<table>
<thead>
<tr>
<th># of Refs</th>
<th>Method</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-train</td>
<td>33.6</td>
</tr>
<tr>
<td>1</td>
<td>First</td>
<td>34.5</td>
</tr>
<tr>
<td>1</td>
<td>Shuffle</td>
<td>35.2</td>
</tr>
<tr>
<td>1</td>
<td>Uniform</td>
<td>36.0</td>
</tr>
</tbody>
</table>

- pre-train on 1M NIST
- fine-tune on 4,667 NIST02-05
- test results on 691 NIST08

Learning curve of different methods with 50 references
## NIST Machine Translation

<table>
<thead>
<tr>
<th># of Refs</th>
<th>Method</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-train</td>
<td>fine-tune</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Pre-train</td>
<td>33.6</td>
</tr>
<tr>
<td>1</td>
<td>First</td>
<td>34.5</td>
</tr>
<tr>
<td>1</td>
<td>Shuffle</td>
<td>35.2</td>
</tr>
<tr>
<td>1</td>
<td>Uniform</td>
<td>36.0</td>
</tr>
</tbody>
</table>

- **pre-train on 1M NIST**
- **fine-tune on 4,667 NIST02-05**
- **test results on 691 NIST08**

Learning curve of different methods with 50 references:

- Pre-train
- Uniform
- Shuffle Sample

+2.4
### Results on Test set

#### NIST Machine Translation

<table>
<thead>
<tr>
<th># of Refs</th>
<th>Method</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-train</td>
<td>1</td>
<td>33.6</td>
</tr>
<tr>
<td>fine-tune</td>
<td>1</td>
<td>34.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>35.2</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>36.0</td>
</tr>
</tbody>
</table>

- pre-train on 1M NIST
- fine-tune on 4,667 NIST02-05
- test results on 691 NIST08

#### MSCOCO Image Captioning

<table>
<thead>
<tr>
<th># of Refs</th>
<th>Method</th>
<th>BLEU</th>
<th>CIDEr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First</td>
<td>26.70</td>
<td>80.70</td>
</tr>
<tr>
<td></td>
<td>Sample One</td>
<td>28.67</td>
<td>85.41</td>
</tr>
<tr>
<td></td>
<td>Shuffle</td>
<td>30.94</td>
<td>94.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+2.3)</td>
<td>(+8.7)</td>
</tr>
<tr>
<td>20</td>
<td>Shuffle</td>
<td>31.79</td>
<td>97.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(+3.1)</td>
<td>(+11.7)</td>
</tr>
</tbody>
</table>

MSCOCO Image Captioning test results
Conclusions

• We use multiple-reference for training MT (image captioning)
• We introduce a soft alignment based lattice compression framework
  • which can generate more training references based on existing ones
  • our methods outperforms baselines on both MT and image captioning
Conclusions

• We use multiple-reference for training MT (image captioning)

• We introduce a soft alignment based lattice compression framework

• which can generate more training references based on existing ones

• our methods outperforms baselines on both MT and image captioning

We’re hiring PhDs

Thanks!

We’re hiring researchers

Codes available at: https://github.com/renj/pseudo-ref