MT/IE: Cross-lingual Open Information Extraction with Neural Sequence-to-Sequence Models

Sheng Zhang, Kevin Duh and Benjamin Van Durme
Johns Hopkins University

EACL | Valencia, April 5th, 2017
Chris wants to build a boat.
Chris wants to build a boat.
Chris wants to build a boat.
Chris wants to build a boat.

Source

Translation

Target

Cross-lingual Open IE

(Chris, wants, Chris build a boat)

(Chris, build, a boat)
Chris wants to build a boat.
Chris wants to build a boat.

Cross-lingual Open IE

Fluency ↑ Adequacy
Cross-lingual Open IE

Solution: Encoder-Decoder Architecture

Sentence in the source language
“克里斯想造一艘船。”

Encoder

(Chris, wants, Chris build a boat)
(Chris, build, a boat)

Decoder

Facts in the target language
Data Annotation

“克里斯想造一艘船。”

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
Data Annotation

Bitext

“克里斯想造一艘船。”

Chris wants to build a boat.

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
Data Annotation

“克里斯 想 造 一 艘 船。”

Chris wants to build a boat.

Universal Dependencies
(http://universaldependencies.org/)

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
Data Annotation

Bitext

“克里斯 想 造 一 艘 船。”

Chris wants to build a boat.

SyntaxNet

Universal Dependencies
(http://universaldependencies.org/)

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
“克里斯 想 造 一 艘 船。”
Chris wants to build a boat.
(Chris, wants, Chris build a boat)
(Chris, build, a boat)
Data Annotation

Bitext

“克里斯 想 造 一 艘 船。”

Chris wants to build a boat.

SyntaxNet

Universal Dependencies
(http://universaldependencies.org/)

PredPatt (interpretable & extensible patterns)

(Chris, wants, Chris build a boat)
(Chris, build, a boat)

40+ languages
Cross-lingual Open IE

Solution: Encoder-Decoder Architecture

Sentence in the source language
“克里斯想造一艘船。”

Encoder

(Chris, wants, Chris build a boat)
(Chris, build, a boat)

Decoder

Facts in the target language
Cross-lingual Open IE

Solution: Encoder-Decoder Architecture

Attention-based Sequence-to-Sequence Model

Sentence in the source language
(e.g., 克里斯 想 造 一艘 船。)

Encoder

Facts in the target language

Decoder
Cross-lingual Open IE

Solution: Encoder-Decoder Architecture

Attention-based Sequence-to-Sequence Model

Sentence in the source language
(e.g., 克里斯 想 造 一 艘 船。)

Encoder

Decoder

Facts in the target language

How to represent facts in a sequence?

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
PredPatt: Predicative Patterns
(https://github.com/hltcoe/PredPatt)

Tuples

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
**PredPatt**: Predicative Patterns

(https://github.com/hltcoe/PredPatt)

Tuples

(Chris, **wants**, Chris build a boat)
(Chris, **build**, a boat)

PredPatt Tree

White et al, EMNLP 2016
**PredPatt**: Predicative Patterns

(https://github.com/hltcoe/PredPatt)

Tuples

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
**PredPatt**: Predicative Patterns

(https://github.com/hltcoe/PredPatt)

Tuples

(Chris, wants, Chris build a boat)
(Chris, build, a boat)

```plaintext
```

PredPatt Tree

White et al, EMNLP 2016
**PredPatt**: Predicative Patterns  
(https://github.com/hltcoe/PredPatt)

Linearized PredPatt

\[
[(\text{Chris} : a_h) \text{ wants} : p_h \text{ SOMETHING } [(\text{Chris} : a_h) \text{ build} : p_h (a : a \text{ boat} : a_h)]]
\]
**PredPatt**: Predicative Patterns

(https://github.com/hltcoe/PredPatt)

Tuples

(Chris, wants, Chris build a boat)
(Chris, build, a boat)

Linearized PredPatt

\[
[(Chris:a_h) \text{ wants}:p_h \text{ something} [(Chris:a_h) \text{ build}:p_h (a:a \text{ boat}:a_h)]]
\]

PredPatt Tree

White et al, EMNLP 2016
**PredPatt:** Predicative Patterns

(https://github.com/hltcoe/PredPatt)

**Tuples**

(Chris, wants, Chris build a boat)
(Chris, build, a boat)

**Linearized PredPatt**

\[
((\text{Chris}:a_h) \ \text{wants}:p_h \ \text{SOMETHING} \ [(\text{Chris}:a_h) \ \text{build}:p_h (a:a \ \text{boat}:a_h)])
\]

**PredPatt Tree**

White et al, EMNLP 2016
Cross-lingual Open IE

⇒ 克里斯 想 造 一艘 船 。

≤ [(Chris:a_h) wants:p_h SOMETHING [(Chris:a_h) build:p_h (a:a boat:a_h)]]
Cross-lingual Open IE

⇒ 克里斯想造一艘船。
Cross-lingual Open IE

⇒ 克里斯想造一艘船。

Chris: a
wants: pₙ
SOMETHING
Chris: aₙ
build: pₙ
(a: a
boat: aₙ)
Cross-lingual Open IE

⇒ 克里斯想造一艘船。

Joint Seq2Seq \rightarrow \text{Neural Seq2Seq} \rightarrow \text{Baseline 1} \rightarrow \text{Joint Moses} \rightarrow \text{Moses}

\[\left[(\text{Chris:} a_h) \text{ wants:} p_h \text{ SOMETHING } [(\text{Chris:} a_h) \text{ build:} p_h (a: a \text{ boat:} a_h)]\right]\]
Cross-lingual Open IE

⇒ 克里斯 想 造 一艘 船。

\[ ((Chris:a_h) \text{ wants}:p_h \text{ SOMETHING} [(Chris:a_h) \text{ build}:p_h (a:a \text{ boat}:a_h)]) \]
Experiment Data

(Chinese, English)
Experiment Data

Bitext: (Chinese, English)
Experiment Data

Bitext: (Chinese, English)

1 million pairs: (Chinese, English Linearized PredPatt)
Experiment Data

Bitext:  (Chinese, English)

1 million pairs: (Chinese, English Linearized PredPatt)

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>950K</td>
<td>10K</td>
<td>40K</td>
</tr>
</tbody>
</table>
Evaluation

Cross-lingual Open IE

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
Evaluation

Cross-lingual Open IE

Translation & Information Extraction

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
Evaluation

Cross-lingual Open IE

Translation & Information Extraction

BLEU \quad F_1

(Chris, wants, Chris build a boat)

(Chris, build, a boat)
Evaluation (BLEU)

PredPatt

[((Chris:a\_h) wants:p\_h SOMETHING [(Chris:a\_h) build:p\_h (a:a boat:a\_h)]])

Predicates

[?arg wants:p\_h SOMETHING [?arg build:p\_h ?arg]]
Evaluation (BLEU)

PredPatt

\[
((Chris:a_h) \text{ wants}:p_h \text{ SOMETHING} ((Chris:a_h) \text{ build}:p_h (a:a \text{ boat}:a_h)))
\]

Predicates

\[
(?\text{arg wants}:p_h \text{ SOMETHING} (?\text{arg build}:p_h \text{ ?arg})]
\]

<table>
<thead>
<tr>
<th></th>
<th>PredPatt</th>
<th>Predicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipeline</td>
<td>17.19</td>
<td>17.24</td>
</tr>
<tr>
<td>Joint Moses</td>
<td>18.34</td>
<td>16.43</td>
</tr>
<tr>
<td>Joint Seq2Seq</td>
<td><strong>18.94</strong></td>
<td><strong>21.55</strong></td>
</tr>
</tbody>
</table>

Table 1: Evaluation results (BLEU)
Evaluation (BLEU)

PredPatt

\[ [(Chris:a_h) \text{ wants:} p_h \text{ SOMETHING } [(Chris:a_h) \text{ build:} p_h (a:a \text{ boat:} a_h)]] \]

Predicates

\[ [\text{?arg wants:} p_h \text{ SOMETHING } [\text{?arg build:} p_h \text{ ?arg}]] \]

<table>
<thead>
<tr>
<th></th>
<th>PredPatt</th>
<th>Predicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipeline</td>
<td>17.19</td>
<td>17.24</td>
</tr>
<tr>
<td>Joint Moses</td>
<td>18.34</td>
<td>16.43</td>
</tr>
<tr>
<td>Joint Seq2Seq</td>
<td>18.94</td>
<td>21.55</td>
</tr>
</tbody>
</table>

Table 1: Evaluation results (BLEU)
Evaluation ($F_1$)

$F_1$ score over predicate tokens

Liu et al, NAACL 2015
Evaluation ($F_1$)

$F_1$ score over predicate tokens

<table>
<thead>
<tr>
<th></th>
<th>$k=1252$</th>
<th>$k=150$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipeline</td>
<td>27.20</td>
<td>28.73</td>
</tr>
<tr>
<td>Joint Moses</td>
<td>25.43</td>
<td>27.94</td>
</tr>
<tr>
<td>Joint Seq2Seq</td>
<td>28.03</td>
<td>29.21</td>
</tr>
</tbody>
</table>

Table 2: Evaluation results ($F_1$ score)
Evaluation ($F_1$)

$F_1$ score over predicate tokens

<table>
<thead>
<tr>
<th>Method</th>
<th>k=1252</th>
<th>k=150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipeline</td>
<td>27.20</td>
<td>28.73</td>
</tr>
<tr>
<td>Joint Moses</td>
<td>25.43</td>
<td>27.94</td>
</tr>
<tr>
<td>Joint Seq2Seq</td>
<td><strong>28.03</strong></td>
<td><strong>29.21</strong></td>
</tr>
</tbody>
</table>

Table 2: Evaluation results ($F_1$ score)

Liu et al, NAACL 2015
Evaluation

Unrecoverable Outputs: \{malformed, empty\}

<table>
<thead>
<tr>
<th>Method</th>
<th>Unrecoverable Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipeline</td>
<td>5615 (15%)</td>
</tr>
<tr>
<td>Joint Moses</td>
<td>33178 (84%)</td>
</tr>
<tr>
<td>Joint Seq2Seq</td>
<td>557 (1%)</td>
</tr>
</tbody>
</table>

Table 3: Number of unrecoverable outputs
### Example

<table>
<thead>
<tr>
<th>Chinese sentence:</th>
<th>重点 审计 关注 与 老百姓 生活 密切 相 关 的 专项 资金.</th>
</tr>
</thead>
<tbody>
<tr>
<td>English sentence:</td>
<td>The focus of the auditing will be on special item funds that are closely related to people’s living.</td>
</tr>
</tbody>
</table>
| Gold             | (1) (The **focus** of the auditing) will be on special item **funds**  
                   (2) (the special item **funds**) are closely **related** to (people ’s **living**) |
| Pipeline:        | (the key auditing **concern** and ordinary people) are closely **related** to (the **life** of the special funds) |
| Joint Moses:     | (the auditing **focus** (**attention**) to (**life**) with (ordinary **people**) are closely **related** to (the special funds)) |
| Joint Seq2Seq:   | (1) (The **focus** of focus) **focused** on (the special **collection** of the specific funds)  
                   (2) (the special **funds**) **related** to (people ’s **lives**) |
Chinese sentence: 重点 审计 关注 与 老百姓 生活 密切 相关 的 专项 资金.

English sentence: The focus of the auditing will be on special item funds that are closely related to people’s living.

Gold
(1) (The focus of the auditing) will be on special item funds
(2) (the special item funds) are closely related to (people ’s living)

Pipeline:
(the key auditing concern and ordinary people) are closely related to (the life of the special funds)

Joint Moses:
(the auditing focus attention to (life) with (ordinary people) are closely related to (the special funds)

Joint Seq2Seq:
(1) (The focus of focus) focused on (the special collection of the specific funds)
(2) (the special funds) related to (people ’s lives)
### Example

<table>
<thead>
<tr>
<th>Chinese sentence:</th>
<th>重点 审计 关注 与 老百姓 生活 密切 相关 的 专项 资金。</th>
</tr>
</thead>
<tbody>
<tr>
<td>English sentence:</td>
<td>The focus of the auditing will be on special item funds that are closely related to people’s living.</td>
</tr>
</tbody>
</table>
| **Gold**         | (1) (The **focus** of the auditing) will be on special item **funds**  
(2) (the special item **funds**) are closely **related** to (people ’s **living**) |
| **Pipeline:**    | (the key auditing **concern** and ordinary people) are closely **related** to (the **life** of the special funds) |
| **Joint Moses:** | (the auditing **focus** (**attention**) to (**life**) with (ordinary **people**) are closely **related** to (the special funds)) |
| **Joint Seq2Seq:** | (1) (The **focus** of focus) **focused** on (the special **collection** of the specific funds)  
(2) (the special **funds**) **related** to (people ’s **lives**)) |
**Example**

| Chinese sentence: | 重点 审计 关注 与 老百姓 生活 密切 相关 的 专项 资金。 |
| English sentence: | The focus of the auditing will be on special item funds that are closely related to people’s living. |
| Gold (1) | (The **focus** of the auditing) will be on special item **funds** |
| | (2) (the special item **funds**) are closely **related** to (people ‘s **living**) |
| Pipeline: | (the key auditing **concern** and ordinary people) are closely **related** to (the **life** of the special funds) |
| Joint **Moses:** | (the auditing **focus** (attention) to (life) with (ordinary **people**) are closely **related** to (the special funds)) |
| **Joint Seq2Seq:** | (1) (The **focus** of focus) **focused** on (the special **collection** of the specific funds) |
| | (2) (the special **funds**) **related** to (people ’s **lives**)) |
**Example**

| Chinese sentence: | 重点 审计 关注 与 老百姓 生活 密切 相关 的 专项 资金。 |
| English sentence: | The focus of the auditing will be on special item funds that are closely related to people’s living. |

| Gold | (1) (The **focus** of the auditing) will be on special item **funds**  
(2) (the special item **funds**) are closely **related** to (people ’s **living**) |
| Pipeline: | (the key auditing **concern** and ordinary people) are closely **related** to (the **life** of the special funds) |
| Joint Moses: | (the auditing **focus** (attention) to (life) with (ordinary **people**) are closely **related** to (the special funds)) |
| Joint Seq2Seq: | (1) (The **focus** of focus) **focused** on (the special **collection** of the specific funds)  
(2) (the special **funds**) **related** to (people ’s **lives**) |
| Chinese sentence: | 重点 审计 关注 与 老百姓 生活 密切 相关 的 专项 资金。 |
| English sentence: | The focus of the auditing will be on special item funds that are closely related to people’s living. |
| Gold | (1) (The **focus** of the auditing) will be on special item **funds**  
(2) (the special item **funds**) are closely **related** to (people ’s **living**) |
| Pipeline: | (the key auditing **concern** and ordinary people) are closely **related** to (the **life** of the special funds) |
| Joint Moses: | (the auditing **focus** (attention) to (life) with (ordinary **people**) are closely **related** to (the **life** of the **funds**)) |
| Joint Seq2Seq: | (1) (The **focus** of focus) **focused** on (the special **collection** of the specific funds)  
(2) (the special **funds**) **related** to (people ’s **lives**) |
Conclusions

Cross-lingual Open IE

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
Conclusions

Encoder \rightarrow Decoder

(Chris, wants, Chris build a boat)
(Chris, build, a boat)
Conclusions

Chris wants to build a boat.

(Chris, wants, Chris build a boat)
(Chris, build, a boat)

Parsing

Extracting

Translating

Chris wants to build a boat.
Conclusions

Encoder ➔ Decoder

(Chris, wants, Chris build a boat)
(Chris, build, a boat)

Parsing

Extracting

Chris wants to build a boat.
Next…

- Adapt the model architecture.
- Apply the framework to low-resource language.
- Use cross-lingual open IE results in downstream tasks.
Cross-lingual Open IE

谢谢大家！→ (Thank, you all)