Adapters for Enhanced Modeling of Multilingual Knowledge and Text

Yifan Hou, Wenxiang Jiao, Meizhen Liu, Carl Allen, Zhaopeng Tu, Mrinmaya Sachan
MLKGs have rich knowledge
MLKGs have rich knowledge
But they are highly incomplete:
- Missing triples / entities / alignments
Multilingual Knowledge Graph + Multilingual Language Model

(MLKG)
- MLKGS have rich knowledge
- But they are highly incomplete:
  - Missing triples / entities / alignments

(MLLM)
- MLLMs have strong transferability
  - Transferring knowledge across 100+ languages
    - Devlin et al., 2019, Conneau et al., 2020
- But they lack factual/multilingual knowledge
  - Pretraining cannot capture much/many:
    - Sparse factual knowledge
    - Features of low resource languages
Multilingual Knowledge Graph + Multilingual Language Model

(MLKG)  ▪ MLKGs have rich knowledge
       ▪ But they are highly incomplete:
           ▪ Missing triples / entities / alignments

(MLLM)  ▪ MLLMs have strong transferability
       ▪ Transferring knowledge across 100+ languages
           ▪ Devlin et al., 2019, Conneau et al., 2020
       ▪ But they lack factual/multilingual knowledge
           ▪ Pretraining cannot capture much/many:
               ▪ Sparse factual knowledge
               ▪ Features of low resource languages

Combining MLKG and MLLM?
▪ MLLM makes MLKG more complete
▪ MLKG makes MLLM more “knowledgeable”
Knowledge Representations

- Multilingual knowledge graph (MLKG)
Knowledge Representations

- Multilingual knowledge graph (MLKG)
  - Factual knowledge triples
  - Cross-lingual entity alignments
**Knowledge Representations**

- **Multilingual knowledge graph (MLKG)**
  - Factual knowledge triples
  - Cross-lingual entity alignments

**Embedding objective:**
- $h =$ head entity
- $r =$ relation
- $t =$ tail entity
- $||h + r - t||$
- TransE (Bordes et al., 2013)
Knowledge Representations

- Multilingual knowledge graph (MLKG)
  - Factual knowledge triples
  - Cross-lingual entity alignments

Knowledge graph embedding

Entity alignment embedding

Embedding objective:
- $h =$ head entity
- $r =$ relation
- $t =$ tail entity
- $||h + r - t||$
- TransE (Bordes et al., 2013)
- $||h - h'||$
Knowledge Representations

- Multilingual knowledge graph (MLKG)
  - Factual knowledge triples
  - Cross-lingual entity alignments

![Diagram of knowledge graph embedding, entity alignment embedding, and MLKG embedding.]

**Embedding objective:**
- \( h = \text{head entity} \)
- \( r = \text{relation} \)
- \( t = \text{tail entity} \)
- \( ||h + r - t|| \)
  - TransE (Bordes et al., 2013)
- \( ||h - h'|| \)
  - \( ||h+r-t|| + ||h-h'|| + ||t-t'|| \)
  - MTransE (Chen et al., 2017)
Knowledge-Aware Multilingual Language Model (MLLM)

- Knowledge in token representations
  - Contextualized representation $t_{\text{Switzerland}}$ should contain:
    \[[t_{\text{Zurich}}, t_{\text{is}}, t_{\text{the}}, t_{\text{largest}}, t_{\text{city}}, t_{\text{in}}, t_{\text{Switzerland}}]\]
Knowledge-Aware Multilingual Language Model (MLLM)

- Knowledge in token representations
  - Contextualized representation $t_{\text{Switzerland}}$ should contain:
  
  - 1. Factual knowledge: (Zurich, is located in, Switzerland)
    - $\text{Average}(t_{\text{Zurich}}, t_{\text{is}}, t_{\text{located}}, t_{\text{in}}) \approx t_{\text{Switzerland}}$
    - KnowBERT (Peters et al., 2019), ERNIE (Zhang et al., 2019)
Knowledge-Aware Multilingual Language Model (MLLM)

- Knowledge in token representations
  - **Contextualized representation** $t_{\text{Switzerland}}$ should contain:
    - 1. Factual knowledge: (Zurich, is located in, **Switzerland**)
      - Average($t_{\text{Zurich}}, t_{\text{is}}, t_{\text{located}}, t_{\text{in}}$) ≈ $t_{\text{Switzerland}}$
        - KnowBERT (Peters et al., 2019), ERNIE (Zhang et al., 2019)
      - $t_{\text{Switzerland}} \approx t_{\text{Svizzera}}$
        - Universal semantic space

\[
[t_{\text{Zurich}}, t_{\text{is}}, t_{\text{the}}, t_{\text{largest}}, t_{\text{city}}, t_{\text{in}}, t_{\text{Switzerland}}]
\]
\[
[t_{\text{Zurich}}, t_{\text{is}}, t_{\text{located}}, t_{\text{in}}]
\]
\[
[t_{\text{Svizzera}}]
\]
Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

<table>
<thead>
<tr>
<th>Task\Knowledge</th>
<th>Multilingual</th>
<th>Factual</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLKG</td>
<td>Adapter-EP</td>
<td>Adapter-TP</td>
</tr>
<tr>
<td>MLLM</td>
<td>Adapter-ES</td>
<td>Adapter-TS</td>
</tr>
</tbody>
</table>
Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

<table>
<thead>
<tr>
<th>Task\Knowledge</th>
<th>Multilingual</th>
<th>Factual</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLKG</td>
<td>Adapter-EP</td>
<td>Adapter-TP</td>
</tr>
<tr>
<td>MLLM</td>
<td>Adapter-ES</td>
<td>Adapter-TS</td>
</tr>
</tbody>
</table>

- Adapter-EP: MLKG entity alignment
  - Wikidata
Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

<table>
<thead>
<tr>
<th>Task\Knowledge</th>
<th>Multilingual</th>
<th>Factual</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLKG</td>
<td>Adapter-EP</td>
<td>Adapter-TP</td>
</tr>
<tr>
<td>MLLM</td>
<td>Adapter-ES</td>
<td>Adapter-TS</td>
</tr>
</tbody>
</table>

- Adapter-EP: MLKG entity alignment
  - Wikidata
- Adapter-TP: MLKG knowledge triples
  - Wikidata
Knowledge Enhancement with Adapters

Knowledgeable adapter set:
- E/T => entity alignment / knowledge triple
- P/S => phrase-level / sentence-level

<table>
<thead>
<tr>
<th>Task/Knowledge</th>
<th>Multilingual</th>
<th>Factual</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLKG</td>
<td>Adapter-EP</td>
<td>Adapter-TP</td>
</tr>
<tr>
<td>MLLM</td>
<td>Adapter-ES</td>
<td>Adapter-TS</td>
</tr>
</tbody>
</table>

- Adapter-EP: MLKG entity alignment
  - Wikidata
- Adapter-TP: MLKG knowledge triples
  - Wikidata
- Adapter-ES: Knowledge enhancement corpus
  - Wikipedia entity description

Phrase-Level (MLKG)
De Botton (zh)  Zurigo (it)
write, Status Anxiety  is located in, Switzerland

Sentence-Level (MLLM)
De Botton spent his early years in Zurich

Adapter Functions
Knowledge Enhancement with Adapters

- Knowledgeable adapter set:
  - E/T => entity alignment / knowledge triple
  - P/S => phrase-level / sentence-level

<table>
<thead>
<tr>
<th>Task</th>
<th>Knowledge</th>
<th>Multilingual</th>
<th>Factual</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLKG</td>
<td>Adapter-EP</td>
<td>Adapter-TP</td>
<td></td>
</tr>
<tr>
<td>MLLM</td>
<td>Adapter-ES</td>
<td>Adapter-TS</td>
<td></td>
</tr>
</tbody>
</table>

- Adapter-EP: MLKG entity alignment
  - Wikidata
- Adapter-TP: MLKG knowledge triples
  - Wikidata
- Adapter-ES: Knowledge enhancement corpus
  - Wikipedia entity description
- Adapter-TS: Knowledge enhancement corpus:
  - T-REx

Phrase-Level (MLKG)
- De Botton (zh)
- Zurigo (it)
- De Botton spent his early years in Zurich
- write, Status Anxiety
- is located in, Switzerland

Sentence-Level (MLLM)
- De Botton (zh)
- Zurigo (it)
- De Botton spent his early years in Zurich
- write, Status Anxiety
- is located in, Switzerland

Adapter Functions
Pipeline

- Adapter training (knowledge enhancement)
  - Training objectives: contrastive learning
  - InfoNCE loss (cosine) on MLLM output representations
Pipeline

- Adapter training (knowledge enhancement)
  - Training objectives
    - InfoNCE loss (cosine) on MLLM output representations

- Finetuning whole enhanced MLLM on downstream tasks
  - MLLM, adapters, fusion module
  - Fusion Mechanism: attention aggregation
    - AdapterFusion (Pfeiffer et al., 2021)

- Inference
Knowledge triple completion
- Given head entity label and relation in one language, find the tail entity
  - E.g., Italian *(Zurigo, si trova in, Svizzera)*
    - (Zurich, is located in, Switzerland)

Input head entity and relation (it): Zurigo, si trova in

All candidate tail entity (it): [Svizzera, De Botton, ...]

Match *(Zurigo, si trova in, Svizzera)*
Results: MLKG Completion

- Knowledge triple completion
  - 1. Enhanced MLLMs always outperform base MLLMs
Results: MLKG Completion

- Knowledge triple completion
  1. Enhanced MLLMs always outperform base MLLMs
  2. Comparable to existing baselines
     - Especially for zero-shot languages
     - Existing baselines cannot support
Cross-lingual entity alignment
- Given entity label in **English**, find aligned one in **other language**
  - E.g., Italian *(Zurich, Zurigo)*

**Results: MLKG Completion**

- **Input entity (en):** Zurich
- **All candidate entity (it):** [Zurigo, Svizzera, …]
- **Match:** (Zurich, Zurigo)
### Results: MLKG Completion

- **Cross-lingual entity alignment**
  - 1. Enhanced MLLMs always outperform base MLLMs
    - Especially for zero-shot languages

![Hit@1 score (XLMR) chart](chart.png)
Cross-lingual entity alignment

1. Enhanced MLLMs always outperform base MLLMs
   - Especially for zero-shot languages
2. Much better than previous baselines
   - E.g., JEANS
Results: MLLM Tasks

- **Knowledge enhancement**
  - Knowledge-intensive task
  - Relation classification (Köksal and Özgür, 2020)

![Relation classification (F1 score)](chart.png)
Results: MLLM Tasks

- **Knowledge enhancement**
  - Knowledge-intensive task
    - Relation classification (Köksal and Özgür, 2020)
  - General language modelling tasks
    - Question Answering (SQuAD & XQuAD)

![Diagram showing results for relation classification and QA tasks](image)

- **Relation classification (F1 score)**
  - mBERT
  - MTMB
  - Enhanced mBERT

- **QA (EM)**
  - mBERT
  - XLMR

<table>
<thead>
<tr>
<th>Base version</th>
<th>Enhanced version</th>
</tr>
</thead>
<tbody>
<tr>
<td>mBERT</td>
<td>47</td>
</tr>
<tr>
<td>MTMB</td>
<td>52</td>
</tr>
<tr>
<td>Enhanced mBERT</td>
<td>57</td>
</tr>
<tr>
<td>mBERT</td>
<td>52</td>
</tr>
<tr>
<td>XLMR</td>
<td>62</td>
</tr>
<tr>
<td>Base version</td>
<td>47</td>
</tr>
<tr>
<td>Enhanced version</td>
<td>62</td>
</tr>
</tbody>
</table>
Results: MLLM Tasks

- Knowledge enhancement
  - Knowledge-intensive task
    - Relation classification (Köksal and Özgür, 2020)
  - General language modelling tasks
    - Question Answering (SQuAD & XQuAD)
    - Name Entity Recognition (WikiAnn)

<table>
<thead>
<tr>
<th>Relation classification</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>mBERT</td>
<td>56</td>
</tr>
<tr>
<td>MTMB</td>
<td>58</td>
</tr>
<tr>
<td>Enhanced mBERT</td>
<td>62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QA (EM)</th>
<th>mBERT</th>
<th>XLMR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>47</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>57</td>
<td>62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NER (F1)</th>
<th>mBERT</th>
<th>XLMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base version</td>
<td>60</td>
<td>61</td>
</tr>
<tr>
<td>Enhanced version</td>
<td>62</td>
<td>63</td>
</tr>
</tbody>
</table>
Takeaways

1. Combining MLKG and MLLM benefit modeling of both multilingual knowledge and text
   - MLKGs become more complete
   - MLLMs become more knowledgeable

2. Enhancement with adapters and contrastive learning works good

All trained adapters are now available on AdapterHub

Code, datasets, and extended benchmarks

Thanks!