DeepStruct: Pretraining of Language Models for Structure Prediction
ACL 2022
Chenguang Wang*, Xiao Liu*, Zui Chen*, Haoyun Hong, Jie Tang, Dawn Song
Structure prediction is important

Structure prediction has a wide range of applications in NLP area
Structure prediction is important

Structure prediction has a wide range of applications in NLP area

NLP Applications

Search Engine

Knowledge Base

QA System

Dialogue System
Structure prediction is important

Structure prediction has a wide range of applications in NLP area
Structure prediction: Example

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.
Structure prediction: Example

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

*Named entity recognition (NER)*

Born in 1951 in Tbilisi, Iago is a Georgian artist.  

Structure prediction tasks could have flexible output formats
Structure prediction: Example

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

*Named entity recognition (NER)*

Born in 1951 in *Tbilisi, Iago* is a Georgian artist.  

*Joint-entity relation extraction (JER)*

Born in 1951 in *Tbilisi, Iago* is a Georgian artist.

Structure prediction tasks could have flexible output formats
Structure prediction: Example

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

**Named entity recognition (NER)**
Born in 1951 in Tbilisi, Iago is a Georgian artist.

**Joint-entity relation extraction (JER)**
Born in 1951 in Tbilisi, Iago is a Georgian artist.

**Open information extraction (OIE)**
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Structure prediction tasks could have flexible output formats
Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding

Structural Understanding
Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Next Word Prediction

Born in 1951 in Tbilisi, Iago is a Georgian artist.
Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding 🔄

Structural Understanding 🔄

Next Word Prediction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Predict single words
Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Next Word Prediction
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Joint-entity relation extraction
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Predict single words
Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Next Word Prediction
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Joint-entity relation extraction
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Predict single words
Predict structures
Traditional Understanding v.s. Structural Understanding

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Traditional Understanding

Structural Understanding

Next Word Prediction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

city_of_birth

city

person

Predict single words

Predict structures

Structural understanding can be more difficult than traditional understanding
Why is structural understanding challenging for LMs?

Born in 1951 in Tbilisi, Iago is a Georgian artist.
Why is structural understanding challenging for LMs?

Challenge 1: Representation for structure

Born in 1951 in Tbilisi, Iago is a Georgian artist.
Why is structural understanding challenging for LMs?

Challenge 1: Representation for structure

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Named entity recognition

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Open information extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.
Why is structural understanding challenging for LMs?

Challenge 1: Representation for structure

Challenge 2: Unifying different structure prediction tasks

Born in 1951 in Tbilisi, Iago is a Georgian artist.
Born in 1951 in Tbilisi, Iago is a Georgian artist.
DeepStruct: Produce triples from text

Born in 1951 in Tbilisi, Iago is a Georgian artist.
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

- city_of_birth
- city
- person
Born in 1951 in Tbilisi, Iago is a Georgian artist.
DeepStruct: Produce triples from text

**Named entity recognition**

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Born in 1951 in Tbilisi, Iago is a Georgian artist.

**Joint-entity relation extraction**

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Born in 1951 in Tbilisi, Iago is a Georgian artist.
DeepStruct: Produce triples from text

Born in 1951 in Tbilisi, Iago is a Georgian artist.
DeepStruct: Produce triples from text

Structure representation formulated as text-to-triple generation problem for LM

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Named entity recognition
- Iago
- 1951
- Tbilisi
- Georgian artist

Joint-entity relation extraction
- date_of_birth
- city_of_birth

Open information extraction

Triples
DeepStruct: Produce triples from text

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Structure representation formulated as text-to-triple generation problem for LM
DeepStruct: Format of output triples

Joint-entity relation extraction

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Named entity recognition

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Structure representation formulated as text-to-triple generation problem for LM
DeepStruct: Format of output triples

Joint-entity relation extraction
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Named entity recognition
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Open information extraction
Born in 1951 in Tbilisi, Iago is a Georgian artist.

Structure representation formulated as text-to-triple generation problem for LM
DeepStruct: Format of output triples

*Joint-entity relation extraction*

Born in 1951 in Tbilisi, Iago is a Georgian artist.

- (Iago; instance of; person)
- (Tbilisi; instance of; city)
- (Iago; city_of_birth; Tbilisi)

*Named entity recognition*

Born in 1951 in Tbilisi, Iago is a Georgian artist.

- (Iago; instance of; person)
- (Tbilisi; instance of; city)

*Open information extraction*

Born in 1951 in Tbilisi, Iago is a Georgian artist.

- (Iago; is a; Georgian artist)

Structure representation formulated as text-to-triple generation problem for LM
DeepStruct: Produce triples from text

Structure representation formulated as text-to-triple generation problem for LM
DeepStruct: Training

Task: *Joint-entity relation extraction*

Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.

Desired Output: (Iago, city_of_birth, Tbilisi), …

DeepStruct concats input text and structure triple for autoregressive training
DeepStruct: Training

Task: *Joint-entity relation extraction*
Input: Born in 1951 in Tbilisi, Iago is a Georgian artist.
Desired Output: (Iago, city_of_birth, Tbilisi), …

DeepStruct concats input text and structure triple for autoregressive training
DeepStruct: Produce triples from text

Born in 1951 in Tbilisi, Iago is a Georgian artist.

Structure representation formulated as text-to-triple generation problem for LM
DeepStruct: Training data

Task-agnostic Datasets

Multi-task Datasets

DeepStruct could incorporate both task-agnostic and multi-task data
DeepStruct: Training data

Task-agnostic Datasets

Multi-task Datasets

Language Model

DeepStruct could incorporate both task-agnostic and multi-task data

\[ \text{jer: Born in 1951 in Tbilisi, Iago} \quad (\text{Iago, city_of_birth, Tbilisi}) \quad \lt e \gt \]

\[ \lt s \gt \text{jer: Born in 1951 in Tbilisi, Iago} \quad (\text{Iago, city_of_birth, Tbilisi})  \quad \lt e \gt \]
## DeepStruct: Task-agnostic datasets

### Dataset Source
- 6 publicly available datasets:
  - T-REx
  - TEKGEN
  - KELM
  - WebNLG
  - ConceptNet
  - OPIEC

### Dataset Statistics
- ~51M sentences
- ~134M entities
- ~114M relations (triples)

DeepStruct is trained on a large task-agnostic corpus
DeepStruct: Training data

Task-agnostic Datasets

Multi-task Datasets

DeepStruct could incorporate both task-agnostic and multi-task data
DeepStruct: Training data

Task-agnostic Datasets

Multi-task Datasets

Language Model

<\textit{s}> jer: Born in 1951 in Tbilisi, Iago \ldots (Iago, city\textunderscore of\textunderscore birth, Tbilisi) \ldots <e>

DeepStruct could incorporate both task-agnostic and multi-task data
DeepStruct: Multi-task datasets

<table>
<thead>
<tr>
<th>28 Datasets</th>
<th>10 Tasks</th>
<th>~ 700K sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Extraction</td>
<td>Semantic Role Labeling</td>
<td>Coreference Resolution</td>
</tr>
<tr>
<td>ACE2005 Trigger Id</td>
<td>CoNLL05 WSJ</td>
<td>CoNLL12</td>
</tr>
<tr>
<td>ACE2005 Trigger Cl</td>
<td>CoNLL05 Brown</td>
<td>TACRED</td>
</tr>
<tr>
<td>ACE2005 Argument Id</td>
<td>CoNLL12</td>
<td>FewRel 1.0</td>
</tr>
<tr>
<td>ACE2005 Argument Cl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open Information Extraction</td>
<td>Joint Entity and Relation Extraction</td>
<td>Dialogue State Tracking</td>
</tr>
<tr>
<td>OIE2016</td>
<td>CoNLL04</td>
<td>MultiWOZ 2.1</td>
</tr>
<tr>
<td>WEB</td>
<td>ADE</td>
<td></td>
</tr>
<tr>
<td>NYT</td>
<td>NYT</td>
<td>ATIS</td>
</tr>
<tr>
<td>PENN</td>
<td>ACE2005</td>
<td>SNIPS</td>
</tr>
<tr>
<td>Named Entity Recognition</td>
<td>Factual Probe</td>
<td></td>
</tr>
<tr>
<td>OntoNotes</td>
<td>Google-RE</td>
<td></td>
</tr>
<tr>
<td>CoNLL03</td>
<td>T-REx</td>
<td></td>
</tr>
<tr>
<td>GENIA</td>
<td>ACE2005</td>
<td></td>
</tr>
</tbody>
</table>

DeepStruct supports a wide range of downstream applications
Result: DeepStruct 10B vs GPT-3 175B

Average Entity F1 on CoNLL04 and ADE (%)

GPT-3 175B
DeepStruct 10B zero-shot
DeepStruct 10B multi-task
DeepStruct 10B multi-task finetune

ours

DeepStruct 10B model remarkably outperforms GPT-3 175B model

+70.7%
DeepStruct achieved state-of-the-art result on 21 of 28 datasets over 10 tasks.
Scaling Effect

DeepStruct multi-task
DeepStruct multi-task finetune

Larger model further improves DeepStruct performance
Conclusion

DeepStruct: train LM to produce triples from text

DeepStruct 10B zero-shot model largely outperforms GPT-3 175B

State-of-the-art on 21 of 28 datasets over 10 tasks

Code: https://github.com/cgraywang/deepstruct

Thank you for your time and interest!