Zero-Shot Information Extraction as a Unified Text-to-Triple Translation

EMNLP 2021

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, Dawn Song
Information Extraction (IE): extract structures from unstructured data

Information extraction is crucial to many NLP applications
We need a unified information extraction approach

Task-specific Predictions

(head entity, relation, tail entity)  relation  (head entity, is, type)  tail entity  tail entity

Task-specific Models

Task-specific Inputs

Dataset1 Dataset2  Dataset1 Dataset2  Dataset1 Dataset2  Dataset1 Dataset2  Dataset1 Dataset2

IE Tasks

OIE  Relation Classification  NER  Factual Probe  Slot Filling

There are many IE tasks, with different task-specific pipelines
We need a unified information extraction approach

The main issue of existing IE methods: limited transferability
Our approach: a unified framework for information extraction

The basic idea: treat every information extraction problem as a “text-to-triple” problem, i.e., translating input text to output triples
Our method: text-to-triple translation

Same text-to-triple translation is shared across tasks, the only difference is the input encoding.
An open information extraction (OIE) example

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

Input Text

OIE

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

Text-to-Triple Translation

1. Generating

(Fisher, born in, Glasgow)

2. Ranking

(Fisher, is a graduate of, London Opera Centre)

Output Triples

Task Predictions

OIE

(Fisher, born in, Glasgow)

(Fisher, is a graduate of, London Opera Centre)
An OIE example: input and output format

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

OIE
Born in Glasgow, Fisher is a graduate of the London Opera Centre.

1. Generating
2. Ranking

Text-to-Triple Translation

(Fisher, born in, Glasgow)
(Fisher, is a graduate of, London Opera Centre)

Output Triples

Task Predictions

OIE
(Fisher, born in, Glasgow)
(Fisher, is a graduate of, London Opera Centre)
An OIE example: input and output format

Input

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

Encode task priors

Born in Glasgow\textsubscript{NP}, Fisher\textsubscript{NP} is a graduate of the London Opera Centre\textsubscript{NP}

Output

(Fisher; Born in; Glasgow)
(Fisher; is a graduate of; London Opera Centre)

Input and output are designed in a format that is appropriate for OIE
An OIE example: zero-shot translation between input text and output triples

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

1. Generating

Text-to-Triple Translation

(Fisher, born in, Glasgow)
(Fisher, is a graduate of, London Opera Centre)

2. Ranking

The common translation module for all tasks is the key

By leveraging the task priors encoded in the input, we enable the zero-shot transfer of the general knowledge that a pre-trained language model has about the task.
An OIE example: generating triples from input text

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

(OIE)

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

1. Generating

2. Ranking

(Fisher, born in, Glasgow)

(Fisher, is a graduate of, London Opera Centre)

Output Triples

Task Predictions
An OIE example: generating triples from input text

OIE Formulation: Extract a set of sequences from input that are relevant to an argument pair

Input text with encoded task priors: Born in Glasgow, Fisher is a graduate of the London Opera Centre.

The generating stage produces general information about the task via pre-trained language models.
An OIE example: generating triples from input text

Beam search with language model attention weights, beam size=1

Input text with encoded task priors: Born in Glasgow, Fisher is a graduate of the London Opera Centre.

Use the attention scores in pre-trained language models to measure the relevance between the sequence and the argument pair.
An OIE example: ranking the generated triples

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

OIE: Born in Glasgow, Fisher is a graduate of the London Opera Centre.

1. Generating

Text-to-Triple Translation

2. Ranking

Output Triples:

(Fisher, born in, Glasgow)

(Fisher, is a graduate of, London Opera Centre)

Task Predictions:

1. Generating

(Fisher, born in, Glasgow)

(Fisher, is a graduate of, London Opera Centre)
An OIE example: ranking the generated triples

The ranking stage finds triples that are of interest to the task via a ranking model pre-trained on a task-agnostic relational corpus.

Task-agnostic Contrastive Pre-training

- Task-agnostic Corpora
  - Triple $t_p$ in $s$
  - Triple $t_n$ not in $s$
  - Sentence $s$

BERT

Ranking Model

- $t_p$ embedding
- $t_n$ embedding
- $s$ embedding

Sequences are relevant not just in relation aspect

Positive pair

Predict which (sentence, triple) pair actually appeared

Negative pair

Finds the triples express the relational information

Top-2

0.2

0.8

0.6

0.3

(Fisher; born in; London Opera Centre)

(Fisher; born in; Glasgow)

(Fisher; is a graduate of; London Opera Centre)

(Fisher; is a graduate of; Glasgow)
An OIE example: decoding task predictions from output triples

Input Text

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

OIE

Born in Glasgow, Fisher is a graduate of the London Opera Centre.

Text-to-Triple Translation

(Fisher, born in, Glasgow)

2. Ranking

(Fisher, is a graduate of, London Opera Centre)

Output Triples

Task Predictions

1. Generating

(Fisher, born in, Glasgow)

(Fisher, is a graduate of, London Opera Centre)
An OIE example: decoding task predictions from output triples

Output Triples
(Fisher; Born in; Glasgow)
(Fisher; is a graduate of; London Opera Centre)

Task Predictions
(Fisher; Born in; Glasgow)
(Fisher; is a graduate of; London Opera Centre)

The framework encodes task priors in the input text and decodes the output triples to finally produce task predictions.
All information extraction tasks in the same framework

Our DeepEx: Zero-shot IE as Text-to-Triple Translation

Relation Classification

Born in place_of_birth Glasgow, Fisher Gold is a graduate of the London Opera Centre.

Factual Probe

Born in place_of_birth Glasgow NP, Fisher Gold, Gold is a graduate of the London Opera Centre NP.

place_of_birth

Glasgow

The framework encodes task priors in the input text and decodes the output triples to finally produce task predictions.
Results: all three information extraction tasks

Our unified approach achieves state-of-the-art or competitive results on all tasks.
Results: comparison between zero-shot (ours) and supervised performance

Our zero-shot approach outperforms fully supervised task-specific models on open information extraction and relation classification.
Results: comparison between interpretable (ours) and blackbox results

**Missing relations in sentences**

Sentence: Judges’ lodgings, the house once occupied by former Prime Minister Edward Heath at Salisbury.
Gold Triple: (Edward Heath; place_of_death; Salisbury)

**Wrong memories of language models**

Sentence: Nick Lucas’s version, released on Brunswick, was a No.
Gold Triple: (Edward Heath; record_label; Brunswick)

Our approach delivers more interpretable results due to enhanced model transparency
Conclusion

- **DeepEx**
  - Unified framework that solves information extraction tasks
  - Competitive and state-of-the-art performance compared to fully supervised methods
  - Better interpretability through enhanced model transparency
  - Zero-shot information extraction without the need of any task-specific training set
  - Generalization by transferring the latent knowledge that language models have
Thank you for your time and interest!

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