REBEL: Relation Extraction By End-to-end Language generation

Pere-Lluis Huguet Cabot and Roberto Navigli

huguetcabot@babelscape.com, navigli@diag.uniroma1.it

End to End Relation Extraction

Huguet Cabot and Navigli - REBEL: Relation Extraction By End-to-end Language generation - Findings of EMNLP (2021)

End to End Relation Extraction

Raw text given as input

"This Must Be the Place" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album "Speaking in Tongues".

End to End Relation Extraction

Extract triplets made of (subject entity*, relation type, object entity*).

"This Must Be the Place" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album "Speaking in Tongues".

(This Must Be the Place, performer, Talking Heads)
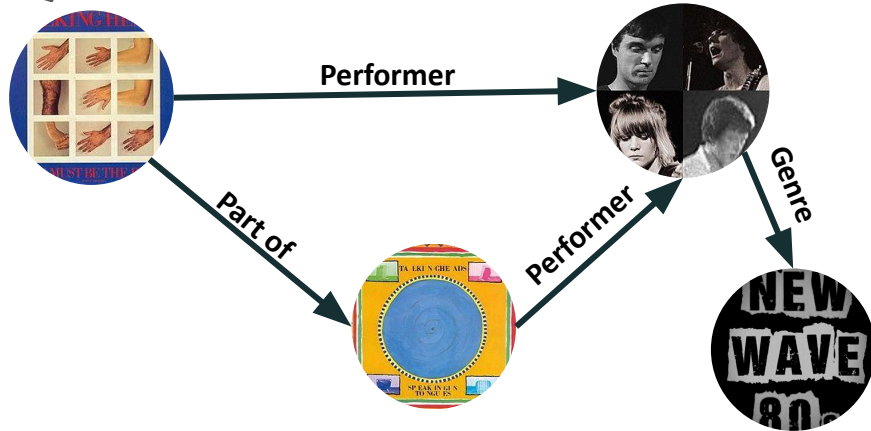(Talking Heads, genre, new wave)
(This Must Be the Place, part of, Speaking in Tongues)
(Speaking in Tongues, performer, Talking Heads)

*For some benchmarks, the entity type also needs to be identified.

REBEL overview

REBEL overview

1 Create a silver dataset 🐊

REBEL overview

1 Create a silver dataset 🐊

2 Linearize triplets

```
<triplet> This Must Be the Place
<subj> Talking Heads <obj> performer
<subj> Speaking in Tongues <obj> part of
<triplet> Talking Heads <subj> new
wave <obj> genre <triplet> Speaking in
Tongues <subj> Talking Heads <obj>
performer
```

REBEL overview

1 Create a silver dataset 🐊

Wikipedia + **WIKIDATA** + 🤗 **NLI**

2 Linearize triplets

<triplet> This Must Be the Place
<subj> Talking Heads <obj> performer
<subj> Speaking in Tongues <obj> part of
<triplet> Talking Heads <subj> new
wave <obj> genre <triplet> Speaking in
Tongues <subj> Talking Heads <obj>
performer

3 Train seq2seq REBEL_{pre-training}

Dataset Creation

Dataset Creation

Leverage Wikipedia and Wikidata

Wikipedia hyperlinks indicate mentions of entities:

This Must Be the Place

From Wikipedia, the free encyclopedia

"**This Must Be the Place**" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album *Speaking in Tongues*.

Dataset Creation

Leverage Wikipedia and Wikidata

Which can be mapped to Wikidata:

Dataset Creation

cRocoDiLe: Automati**c R**elati**o**n Extra**c**ti**o**n **D**ataset w**i**th N**L**I filt**e**ring

- Inspired by T-REx[1].

[1]Hady Elsahar, Pavlos Vougiouklis, Arslen Remaci, Christophe Gravier, Jonathon Hare, Frederique Laforest, and Elena Simperl. 2018. T-REx: A large scale alignment of natural language with knowledge base triples. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association.*

Dataset Creation

cRocoDiLe: Automati**c R**elati**o**n Extra**cti**o**n D**ataset w**i**th NL**I** filt**e**ring

- Inspired by T-REx[1].
- Use Wikipedia abstracts and hyperlinks instead of DBpedia.

[1]Hady Elsahar, Pavlos Vougiouklis, Arslen Remaci, Christophe Gravier, Jonathon Hare, Frederique Laforest, and Elena Simperl. 2018. T-REx: A large scale alignment of natural language with knowledge base triples. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association.*

Dataset Creation

cRocoDiLe: Automati**c** **R**elati**o**n Extra**c**ti**o**n **D**ataset w**i**th N**L**I filt**e**ring

- Inspired by T-REx[1].
- Use Wikipedia abstracts and hyperlinks instead of DBpedia.
- We create a distantly annotated dataset by interlinking Wikipedia and Wikidata.

[1]Hady Elsahar, Pavlos Vougiouklis, Arslen Remaci, Christophe Gravier, Jonathon Hare, Frederique Laforest, and Elena Simperl. 2018. T-REx: A large scale alignment of natural language with knowledge base triples. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association.*

Dataset Creation

cRocoDiLe: Automati**c** **R**elati**o**n Extra**c**ti**o**n **D**ataset w**i**th N**L**I filt**e**ring

- Inspired by T-REx[1].
- Use Wikipedia abstracts and hyperlinks instead of DBpedia.
- We create a distantly annotated dataset by interlinking Wikipedia and Wikidata.
- Overgeneration of non entailed relations:

  Gràcia is a district of the city of Barcelona, Catalonia.

  Triplets: (Gracia, instance of, district), (Gracia, located in the administrative territorial entity, Barcelona), (Barcelona, capital of, Catalonia)

[1]Hady Elsahar, Pavlos Vougiouklis, Arslen Remaci, Christophe Gravier, Jonathon Hare, Frederique Laforest, and Elena Simperl. 2018. T-REx: A large scale alignment of natural language with knowledge base triples. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association.*

Dataset Creation

cRocoDiLe: Automati**c R**elati**o**n Extra**ct**i**o**n **D**ataset w**i**th N**L**I filt**e**ring

- Inspired by T-REx[1].
- Use Wikipedia abstracts and hyperlinks instead of DBpedia.
- We create a distantly annotated dataset by interlinking Wikipedia and Wikidata.
- Overgeneration of non entailed relations:

  Gràcia is a district of the city of Barcelona, Catalonia.

  Triplets: (Gracia, instance of, district), (Gracia, located in the administrative territorial entity, Barcelona), (Barcelona, capital of, Catalonia)

- Use of an NLI system to filter out such triplets.

[1]Hady Elsahar, Pavlos Vougiouklis, Arslen Remaci, Christophe Gravier, Jonathon Hare, Frederique Laforest, and Elena Simperl. 2018. T-REx: A large scale alignment of natural language with knowledge base triples. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association.*

NLI: Natural Language Inference

We use XNLI RoBERTa (joeddav/xlm-roberta-large-xnli) to identify triplets that are not explained:

Gràcia is a district of the city of Barcelona, Catalonia. <sep> Barcelona capital of Catalonia.

We use the entailment prediction and filter out the triplets under a threshold.

Triplet Linearization

Triplet Linearization

Relation Extraction to seq2seq

"This Must Be the Place" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album "Speaking in Tongues"

(**This Must Be the Place**, performer, Talking Heads)
(Talking Heads, genre, new wave)
(**This Must Be the Place**, part of, Speaking in Tongues)
(Speaking in Tongues, performer, Talking Heads)

}

```
<triplet> This Must Be the Place
<subj>
```

Triplet Linearization

Relation Extraction to seq2seq

"This Must Be the Place" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album "Speaking in Tongues"

(**This Must Be the Place**, **performer**, **Talking Heads**)
(Talking Heads, genre, new wave)
(This Must Be the Place, part of, Speaking in Tongues)
(Speaking in Tongues, performer, Talking Heads)

```
<triplet> This Must Be the Place
<subj> Talking Heads <obj>
performer <subj>
```

Triplet Linearization

Relation Extraction to seq2seq

"This Must Be the Place" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album "Speaking in Tongues"

~~(This Must Be the Place, performer, Talking Heads)~~
(Talking Heads, genre, new wave)
(This Must Be the Place, part of, Speaking in Tongues)
(Speaking in Tongues, performer, Talking Heads)

}

```
<triplet> This Must Be the Place
<subj> Talking Heads <obj>
performer <subj> Speaking in
Tongues <obj> part of
```

Triplet Linearization

Relation Extraction to seq2seq

"This Must Be the Place" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album "Speaking in Tongues"

(This Must Be the Place, performer, Talking Heads)
(**Talking Heads**, genre, new wave)
(This Must Be the Place, part of, Speaking in Tongues)
(Speaking in Tongues, performer, Talking Heads)

}

`<triplet>` This Must Be the Place `<subj>` Talking Heads `<obj>` performer `<subj>` Speaking in Tongues `<obj>` part of **`<triplet>`** **Talking Heads** **`<subj>`**

Triplet Linearization

Relation Extraction to seq2seq

"This Must Be the Place" is a song by new wave band Talking Heads, released in November 1983 as the second single from its fifth album "Speaking in Tongues"

~~(This Must Be the Place~~, ~~performer~~, ~~Talking Heads~~)
~~(Talking Heads~~, ~~genre~~, ~~new wave~~)
~~(This Must Be the Place~~, ~~part of~~, ~~Speaking in Tongues~~)
~~(Speaking in Tongues~~, ~~performer~~, ~~Talking Heads~~)

`<triplet>` This Must Be the Place `<subj>` Talking Heads `<obj>` performer `<subj>` Speaking in Tongues `<obj>` part of `<triplet>` Talking Heads `<subj>` **new wave** **`<obj>` genre `<triplet>` Speaking in Tongues `<subj>` Talking Heads `<obj>` performer**

Results

Results

Comparable systems

- **SpERT** (Eberts and Ulges, 2020), based on BERT.

- **Table-sequence** (Wang and Lu, 2020), based on ALBERT.

- **TANL** (Paolini et al., 2021), based on T5.

- **TPLinker** (Wang et al., 2021) based on BERT.

- **REBEL**, ours, based on BART.
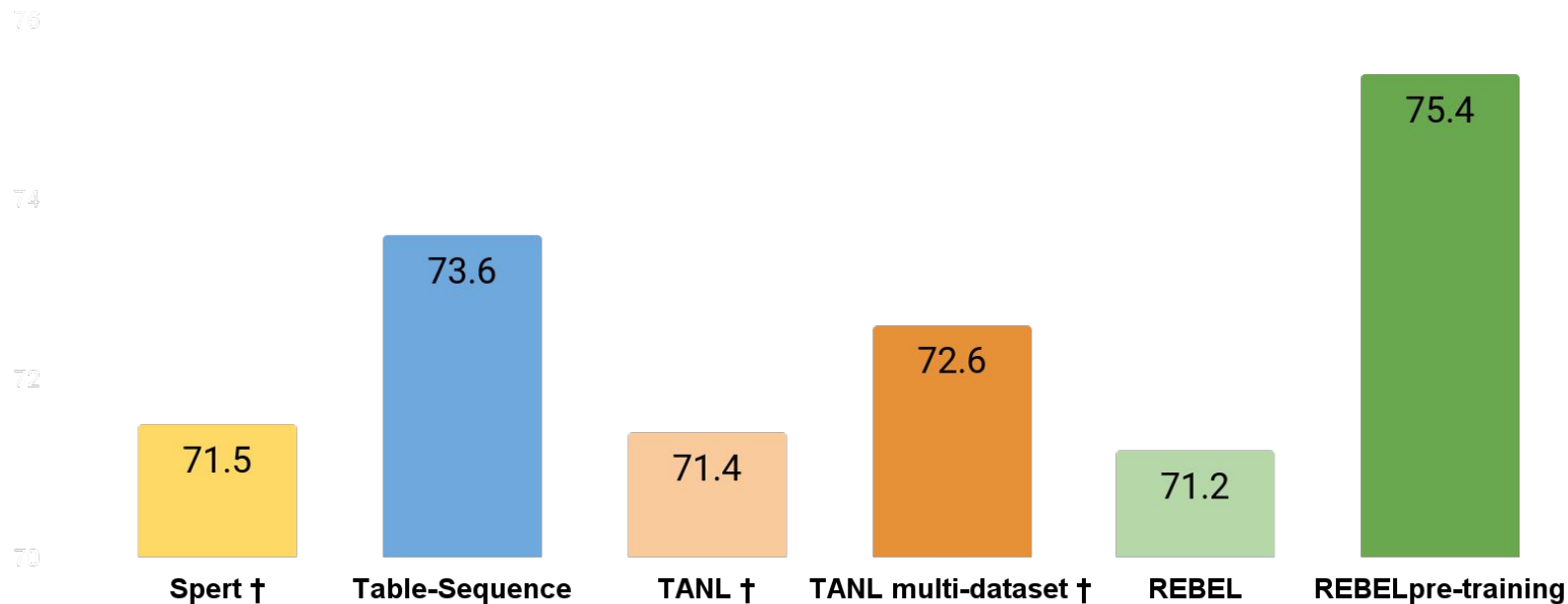
Results

Comparable systems

- **SpERT** (Eberts and Ulges, 2020), based on BERT.

- **Table-sequence** (Wang and Lu, 2020), based on ALBERT.

- **TANL** (Paolini et al., 2021), based on T5.

- **TPLinker** (Wang et al., 2021) based on BERT.

- **REBEL**, ours, based on BART.

Datasets

- **CONLL04** (Roth and Yih, 2004), four entity types, five relation types.

- **ADE** (Gurulingappa et al., 2012), two entity types, one relation type.

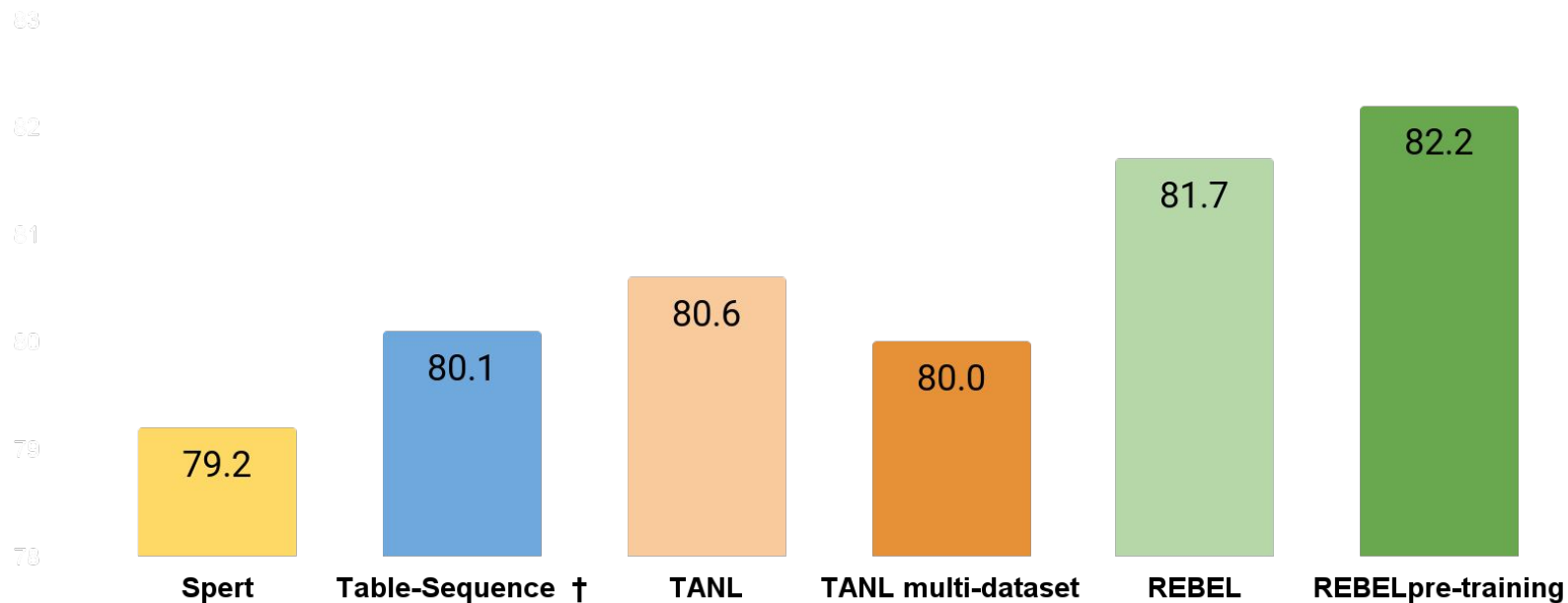- **NYT** (Riedel et al., 2010), three entity types, 24 relation types.

Results

CONLL04 - Micro F1

Spert †	Table-Sequence	TANL †	TANL multi-dataset †	REBEL	REBELpre-training
71.5	73.6	71.4	72.6	71.2	75.4

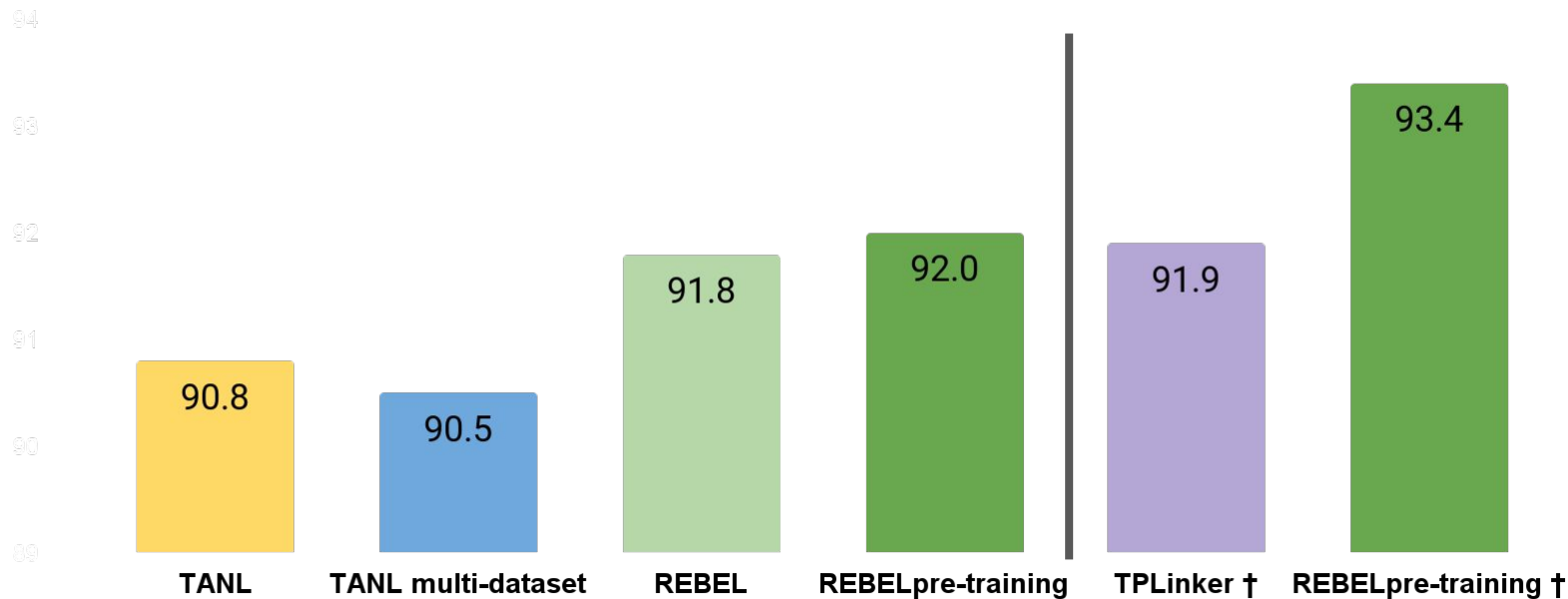† Explicit use of train+dev for training.

Results

ADE - Micro F1

† Filtered overlapping entities.

Results

NYT - Micro F1

† No entity typing.

Thank you for your attention!

Visit **REBEL** at https://github.com/Babelscape/rebel

Babelscape: https://babelscape.com • **Sapienza NLP: http://nlp.uniroma1.it**

@Babelscape   @SapienzaNLP

EMNLP 2021

EU H2020 Project
KnowGraphs No. 860801