Improving Relation Extraction by Pre-trained Language Representations

Christoph Alt*, Marc Hübner*, Leonhard Hennig

https://github.com/DFKI-NLP/TRE

{firstname.lastname}@dfki.de

Summary

Current state-of-the-art relation extraction methods typically rely on **lexical, syntactic, and semantic features, explicitly computed** in a **pre-processing step**, that require additional **annotated language resources**. This severely restricts the **applicability and portability** and introduces a **source of errors**. We introduce **TRE**, a Transformer for Relation Extraction, extending the OpenAI Generative Pre-trained Transformer [Radford et al., 2018]. **TRE** uses pre-trained **deep language representations** instead of explicit linguistic features and allows us to learn **implicit linguistic features** solely from plain text corpora by **unsupervised pre-training**, before fine-tuning the learned language representations on the relation extraction task. **TRE** obtains a **new state-of-the-art** result on the **TACRED** and **SemEval 2010 Task 8** datasets.

Goals

- High performance Relation Extraction
- No task specific architecture
- Limited pre-processing of source corpora
- Limited dependency on domain-specific resources

Challenges

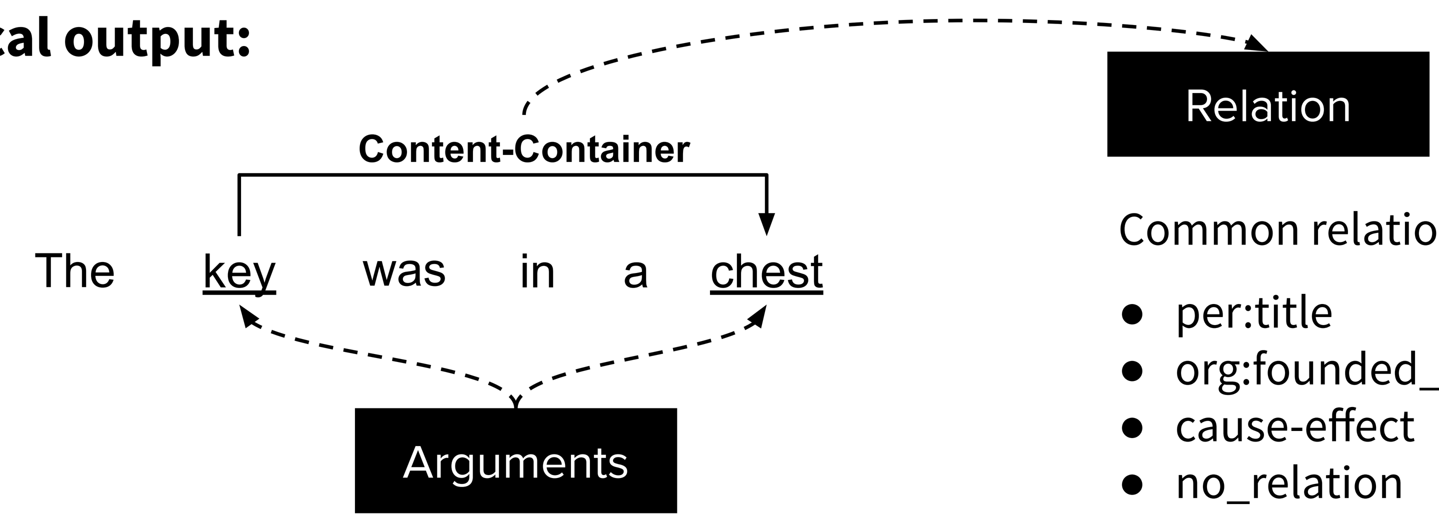
- Entity Masking is still crucial for best performance
- Overfitted language representations
- Hyperparameter tuning?

The Task (Relation Extraction)

The following are typical inputs to a information extraction system:

<u>Mr. Scheider</u> played the <u>police chief</u> of a resort town menaced by a shark.
The measure included <u>Aerolineas</u>'s domestic subsidiary, <u>Austral</u>.
The <u>key</u> was in a <u>chest</u>.

Typical output:



Common relations:

- per:title
- org:founded_by
- cause-effect
- no_relation

Datasets

Dataset	Relation Types	examples	negative examples
TACRED	42	106,264	79.5%
SemEval 2010 Task 8	19	10,717	17.4%

TACRED relation types mostly focus on named entities, whereas SemEval contains semantic relations between concepts.

Results

TACRED

System	P	R	F1
LR†	72.0	47.8	57.5
CNN†	**72.1**	50.3	59.2
Tree-LSTM†	66.0	59.2	62.4
PA-LSTM†	65.7	64.5	65.1
C-GCN†	69.9	63.3	66.4
TRE (ours)	70.1	**65.0**	**67.4**

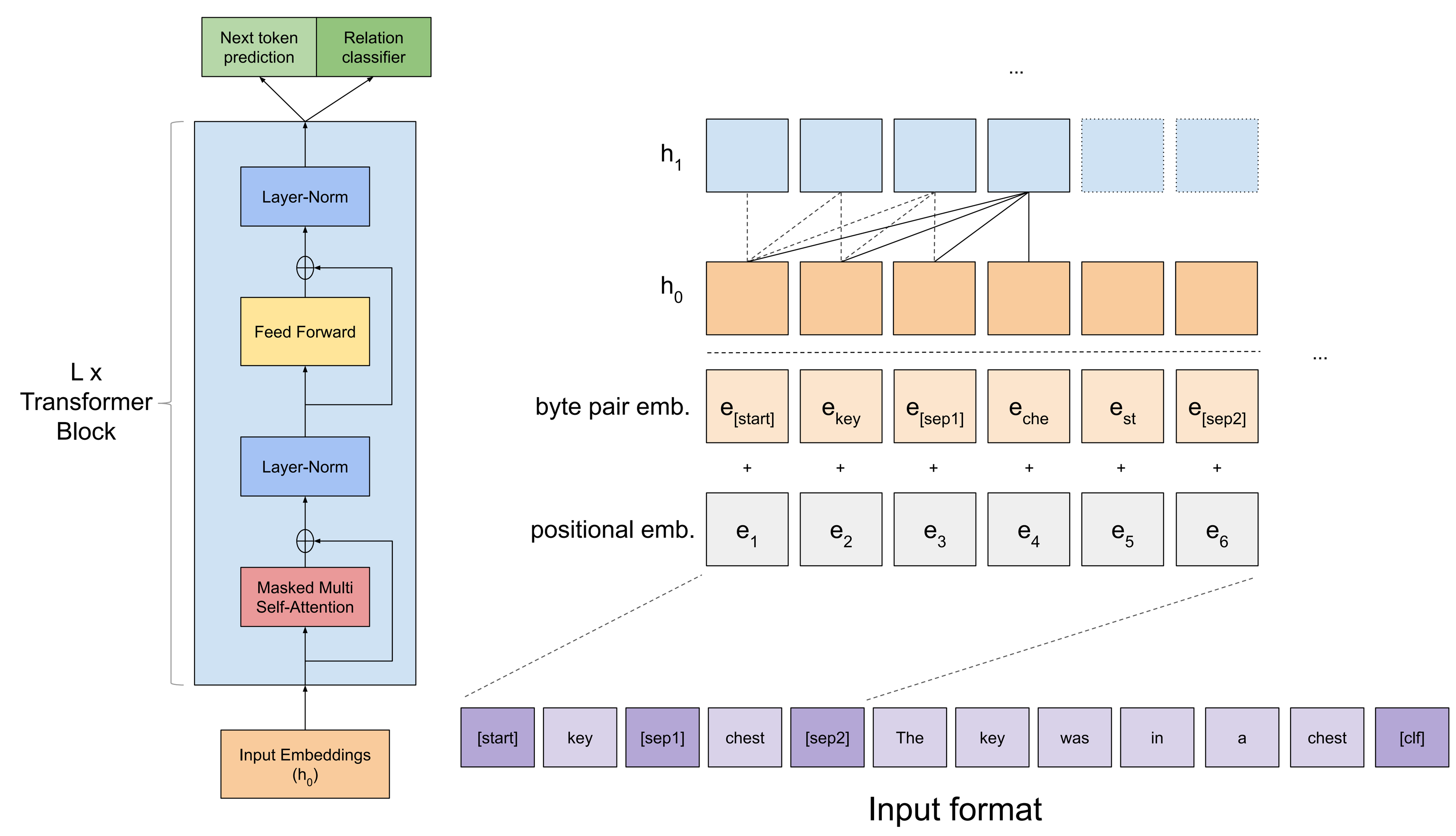
SemEval

System	P	R	F1
SVM†	–	–	82.2
PA-LSTM†	–	–	82.7
C-GCN†	–	–	84.8
DRNN†	–	–	86.1
BRCNN†	–	–	86.3
TRE (ours)	88.0	86.2	**87.1**

† as reported in the original work

TRE Architecture

Fine-tuned Transformer Language Model (OpenAI GPT) on a task-specific input format using multi-task learning with auxiliary loss as regularizer.



Input format

Ablation

	SemEval		TACRED		
	None	UNK	None	UNK	NE + GR
Best model	**85.6**	76.9	63.3	51.0	**68.0**
– w/o pre-trained LM	**75.6**	68.2	43.3	41.6	**64.2**
– w/o pre-trained LM and BPE	55.3	**60.9**	38.5	38.4	**60.8**

Key Findings:

- Entity Masking helps for LM representation generalization
- Without masking LM pre-training learns more generalizable representations for entities of both datasets

Entity Masking on TACRED

None: The measure included <u>Aerolineas</u>'s domestic subsidiary, <u>Austral</u>
UNK: The measure included <u><UNK></u>'s domestic subsidiary, <u><UNK></u>
GR: The measure included <u><SUB></u>'s domestic subsidiary, <u><OBJ></u>
NE: The measure included <u><ORG></u>'s domestic subsidiary, <u><ORG></u>

Entity Masking	Precision	Recall	F1
None	**69.5**	58.1	63.3
UNK	56.9	46.3	51.0
GR	63.8	50.1	56.1
NE	68.8	65.3	67.0
NE + GR	68.8	**67.2**	**68.0**

Key Findings:

- "None" masking delivers high precision but low recall
- Entity type information helps to generalize
- Additional Grammar masking further boosts performance

German Research Center for Artificial Intelligence

DAY STREAM

BERLIN BIG DATA CENTER

Sponsored by:

Federal Ministry of Transport and Digital Infrastructure

Federal Ministry of Education and Research

Supported through projects DEEPLEE (01IW17001) , BBDC2 (01IS18025E), and DAYSTREAM (19F2031A)