Tackling Contradiction Detection in German Using Machine Translation and End-to-End Recurrent Neural Networks

Maren Pielka, Rafet Sifa, Lars Hillebrand, David Biesner,
Rajkumar Ramamurthy, Anna Ladi, and Christian Bauckhage
maren.pielka@iais.fraunhofer.de

Fraunhofer Institute for
Intelligent Analysis and Information Systems IAIS

December 10, 2020

Fraunhofer
IAIS

Agenda

- Introduction to the topic
- Experiments
- Results
- Conclusion and Outlook

Overview

Topic: Exploring methods to detect contradictions in German text
  $\rightarrow$ Sub-topic of Natural Language Inference (NLI)

Data:

- Stanford Natural Language Inference (SNLI) data set,
  partially translated to German by a machine translation engine
  $\rightarrow$ 120 000 labeled sentence pairs

Applications:

- Fact Checking
- Financial Consistency Checks
- Question Answering
- ...

Main task: Finding semantic correlations between two sentences
→ **premise** and **hypothesis**

	Premise	Hypothesis	Label
1	"A person on a horse jumps over a broken down airplane."	"A person is outdoors, on a horse."	"entailment"
2	"A person on a horse jumps over a broken down airplane."	"A person is training his horse for a competition."	"neutral"
3	"A person on a horse jumps over a broken down airplane."	"A person is at a diner, ordering an omelette."	"contradiction"

Table: Example sentence pairs with labels from the SNLI data set

Our contribution:

- First benchmarks in an unexplored area
  - → Most previous work has been done on English text, and for the three-way task (entailment vs. neutral vs. contradiction)
- Applying state-of-the art methods to machine-translated data

Learning Sentence Embeddings

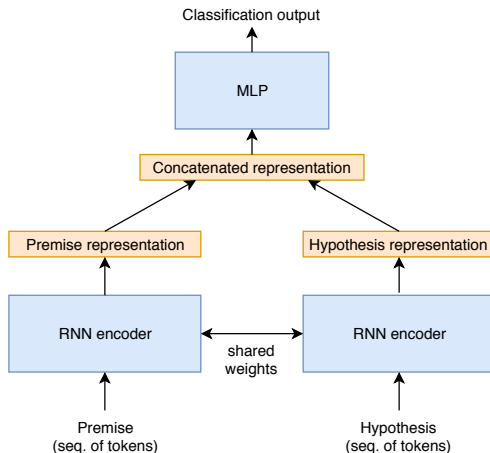**Model 1:** Concatenate embeddings from premise and hypothesis using RNN encoders, and apply an MLP classifier

Figure: RNN model without attention

Learning Sentence Embeddings (cont.)

Model 2: Apply two RNN encoders with an attention mechanism, obtain embeddings from second RNN and apply an MLP classifier
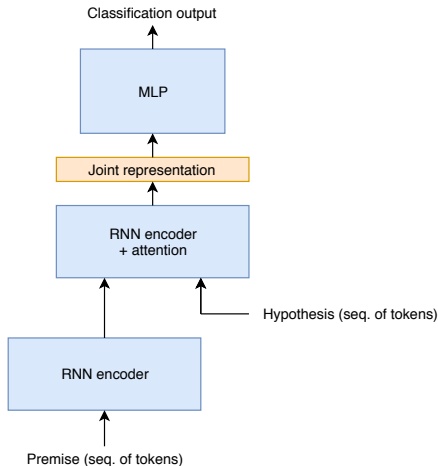

Figure: RNN model with attention

	Original data	**Translated data**
TF-IDF	0.7602	0.7422
Flair (pre-trained, mean-pooled)	0.7703	0.7807
RNN (without attention)	0.7734	**0.7847**
RNN (with attention)	0.7555	0.7807
MBERT	**0.8508**	0.7457

Table: Performance comparison (accuracy) for the different classification methods, evaluated on the original and the translated SNLI test set, respectively.

Results: Attention vs. no attention

	Original sentence pair	Translated sentence pair	Model 1 prediction	Model 2 prediction	True label
1	**P:** "Here is a picture of people getting drunk at a house party." **H:** "People were celebrating at a Christmas themed party."	**P:** "Hier ist ein Bild von Leuten, die sich auf einer Hausparty betrinken." **H:** "Die Leute feierten auf einer Weihnachtsparty."	No contradiction	Contradiction	No contradiction
2	**P:** "A black man in a white uniform makes a spectacular reverse slam dunk to the crowd's amazement." **H:** "the man is Asian"	**P:** "Ein schwarzer Mann in weißer Uniform macht einen spektakulären Reverse-Slam Dunk zum Staunen der Menge." **H:** "der Mann ist Asiate"	No contradiction	No contradiction	Contradiction

Table: Prediction examples for the model without (model 1) and with (model 2) attention, in comparison.

Conclusion

Important findings

- RNN-based models outperform simple methods on both the original and the translated data set.
- Overall, the performances on the original and the translated data set are similar (except for MBERT) , verifying the high quality of the translation.

Future work

- Incorporate world knowledge to the model (e.g. by exploiting knowledge graphs)
- Apply models to real-world data

Thank you for your attention!
Any questions?

References

📄 S. Bowman et al. "A Large Annotated Corpus for Learning Natural Language Inference". *Proc. of EMNLP 2015*. EMNLP. 2015.

📄 A. Conneau et al. "Supervised Learning of Universal Sentence Representations from Natural Language Inference Data". *Proc. of EMNLP 2017*. EMNLP. 2017.

📄 Guillaume Lample and Alexis Conneau. "Cross-lingual Language Model Pretraining". *Advances in Neural Information Processing Systems (NeurIPS)* (2019).

📄 T. Rocktäschel et al. "Reasoning About Entailment with Neural Attention". *Proc. of ICLR 2016*. ICLR. 2016.

📄 Rafet Sifa et al. "Towards Contradiction Detection in German: A Translation-driven Approach". *Proc. of IEEE SSCI 2019*. 2019.