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

Maren Pielka, Rafet Sifa, Lars Patrick Hillebrand, David Biesner, Rajkumar Ramamurthy, Anna Ladi, and Christian Bauckhage

Fraunhofer IAIS

Motivation and Outline

Topic: Exploring methods to detect contradictions in German text, extending the research by [5]
→ Sub-topic of Natural Language Inference (NLI)

Data:

- Stanford Natural Language Inference (SNLI) data set [1], partially translated to German by a machine translation engine → 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

Experimental Setup and Implementation

Our implementation is based on the paradigms from [2] and [4].

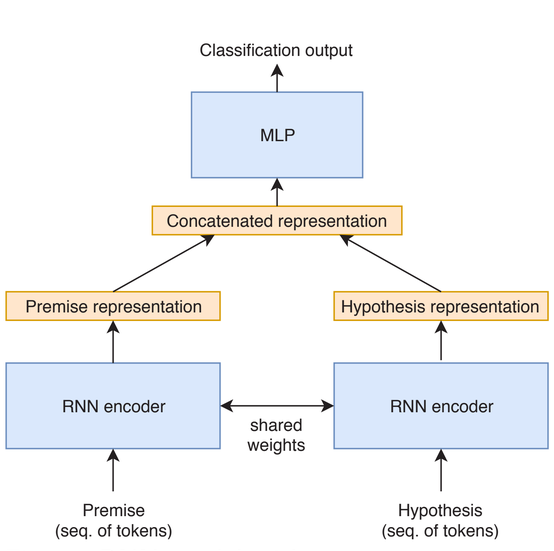

Figure: RNN model without attention, based on [2]

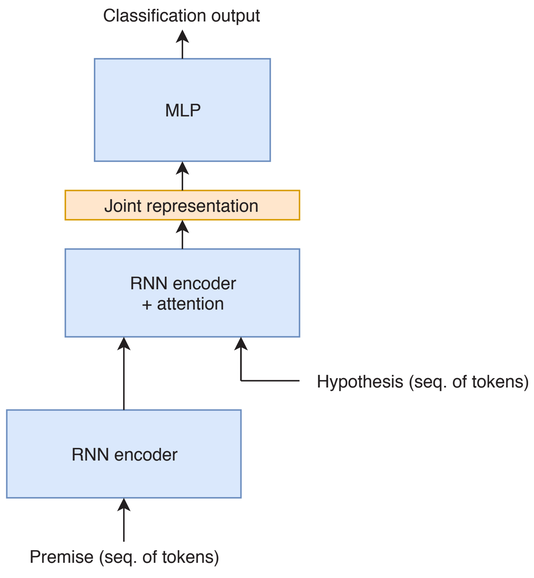

Figure: RNN model with attention, based on [4]

Figure: High-level architecture of our RNN-based models

In addition to our RNN-based models, we utilize a pre-trained MBERT model [3] and fine-tune it on the NLI task for German and English.

Results Overview

	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 Analysis

	Original sentences	Translated sentences	Model 1 prediction	Model 2 prediction	True label
1	**P:** "The man is wearing an orange and black polo shirt and is kneeling with his lunch box in one hand while holding a banana in his other hand." **H:** "A man is wearing a green t shirt while holding a banana."	**P:** "Der Mann trägt ein orange-schwarzes Poloshirt und kniet mit seiner Lunchbox in der einen Hand und hält eine Banane in der anderen Hand." **H:** "Ein Mann trägt ein grünes T-Shirt und hält eine Banane."	No contradiction	Contradiction	Contradiction
2	**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."	Contradiction	No contradiction	No contradiction
3	**P:** "Cityscape with four men on bikes serving as the focal point of picture." **H:** "Local robbers escape after stealing an otter"	**P:** "Stadtlandschaft mit vier Männern auf dem Fahrrad, die als Mittelpunkt des Bildes dient." **H:** "Lokale Räuber fliehen, nachdem sie einen Otter gestohlen haben."	Contradiction	No contradiction	Contradiction
4	**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. The attention-based model can identify contradictions that are linked to specific words (example 1), but it also returns wrong predictions for examples where the alignment between the sentences is not helpful for the classification (examples 2 and 3). Both approaches cannot cope with sentences where world knowledge is involved (example 4).

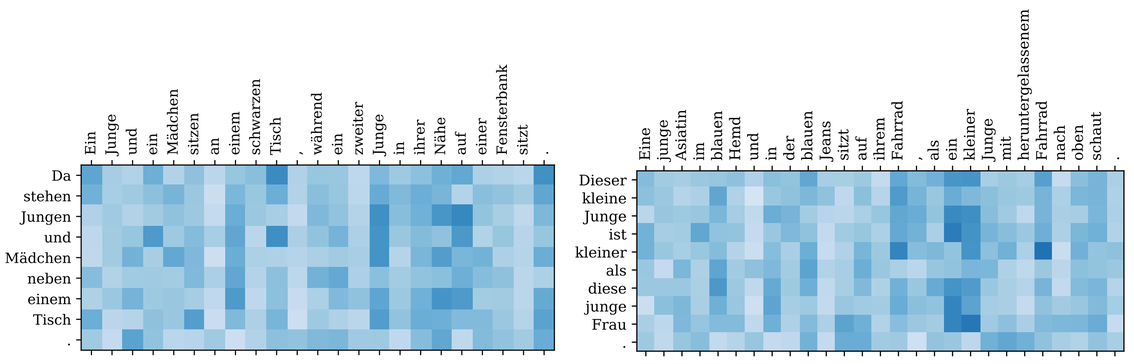

Figure: Attention matrices of the model after inputting two sentence pairs from the translated SNLI test data set. The x-axis corresponds to the premise, and the y-axis corresponds to the hypothesis. The model sometimes finds meaningful alignments, still most of the attention values are rather inconclusive.

Outlook

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

References

[1] S. Bowman, G. Angeli, C. Potts, and C. Manning. A Large Annotated Corpus for Learning Natural Language Inference. In *Proc. of EMNLP 2015*. EMNLP, 2015.

[2] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes. Supervised Learning of Universal Sentence Representations from Natural Language Inference Data. In *Proc. of EMNLP 2017*. EMNLP, 2017.

[3] G. Lample and A. Conneau. Cross-lingual language model pretraining. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[4] T. Rocktäschel, E. Grefenstette, K. Hermann, T. Kocisky, and P. Blunsom. Reasoning About Entailment with Neural Attention. In *Proc. of ICLR 2016*. ICLR, 2016.

[5] R. Sifa, M. Pielka, R. Ramamurthy, A. Ladi, L. Hillebrand, and C. Bauckhage. Towards contradiction detection in german: A translation-driven approach. In *Proc. of IEEE SSCI 2019*, 2019.