Story Cloze Test

Lea Fritschi **Beat Hubmann** **Maya Vögeli** **Lukas Wampfler**

1 Introduction

You can keep this short. Ideally you introduce the task already in a way that highlights the difficulties your method will tackle.

2 Methodology

In our approach we use the idea of **quick thoughts**, as proposed in [1]. The basic idea is to train an encoder in order to learn an embedding for the four context sentences (the first four sentences of the short story). Instead of using a decoder, the model then uses a classifier to identify the correct sentence from the set of two candidate sentences (basically deciding which of the two candidate embeddings is closer to the embedding of the four starting sentences of the story.

Is there a particular additional data source you want to use?

Frage: Wurde anderer Code (zB von [1]) verwendet? Falls ja, sollte man das deklarieren.

3 Model

We use the meaning of the first four sentences to predict the meaning of the last sentence, where meaning is represented by an embedding of the context computed from an encoding function. Our loss function is defined in feature space.

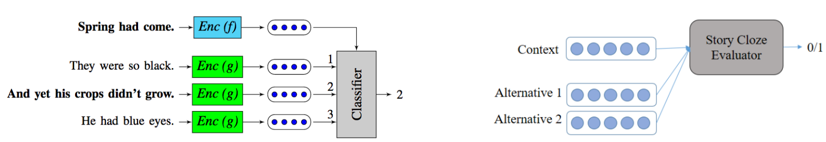
Formally: Let f and g be parametrized functions that take a context of four sentences (f) resp. a single sentence (g) as input and encode it into a fixed length vector of equal size. For a given candidate sentence s_{cand} and the set S_{cand} of all possible candidate sentences (two in our case), the probability that s_{cand} is the correct last sentence, is given by

$$p(s_{cand}|\, cont, S_{cand}) = \frac{exp(c(f(cont), g(s_{cand})}{\sum_{s' \in S_{cand}} exp(c(f(cont), g(s')}$$

where $cont$ denotes the four-sentence context of the last sentence and c is a scoring function/ classifier. **comment: this is taken from [1]: is this correct?**

For the encoder, we used a bidirectional RNN with 600 GRU-cells.

Attention mechanism ? Gibt es dazu etwas zu sagen? **Image of architecture:**



(a) Figure taken from [1]

(b) Figure taken from [2]

Figure 1: **Frage:** Welche Art von Grafik ist für den Report geeigneter (links: müsste ich noch anpassen auf unsere Methode; die Grafik rechts könnte man so übernehmen.

4 Training

For training, we used a corpus of 88161 five sentence stories. In these stories, there were no alternative (wrong) ending sentences. We used GloVe pretrained word embeddings in 300 dimensions.[3].

The training objective maximizes the probability .... (to be continued)

What is your objective? How do you optimize it?

5 Experiments

On the validation set, we achieve an accuracy of 0.60. This result is comparable with respect to what has been published earlier [] This **must** at least include the accuracy of your method on the validation set.

6 Conclusion

You can keep this short, too.

References

[1] L. Logeswaran and H. Lee. An efficient framework for learning sentence representations. *CoRR*, abs/1803.02893, 2018. URL `http://arxiv.org/abs/1803.02893`.

[2] N. Mostafazadeh, L. Vanderwende, W.-t. Yih, P. Kohli, and J. Allen. Story cloze evaluator: Vector space representation evaluation by predicting what happens next. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP*, pages 24–29, Berlin, Germany, August 2016. Association for Computational Linguistics. URL `http://anthology.aclweb.org/W16-2505`.

[3] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014. URL `http://www.aclweb.org/anthology/D14-1162`.