Story Cloze Test

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

1   Introduction

The given task was to design and implement an attempt at the 'Story Cloze Test' (zu zitieren: https://arxiv.org/abs/1604.01696), a framework for evaluating story understanding and script learning. After some disappointing initial attempts with an attention-augmented encoder-decoder model, we chose to apply the idea to learn sentence vector representations to predict a sentence's context [1] to the 'Story Cloze Test'.

2   Methodology

In our implementation, we follow the approach of **quick thoughts**, as proposed in [1]. Unlike the approach of generating a target sequence from an input sentence commonly used for this class of problems, our chosen approach attempts to choose the correct target sentence of an input sentence from a set of given candidate sentences based on learned sentence encodings. A massive benefit of this approach is that it allows for a large (i.e. rich) vocabulary size as no target sequence needs to be generated from the encoded input and there thus is no efficiency and numerical stability bottleneck at that point. Also, this approach uses the very simple and efficient classifier $c(u, v) = u^T v$ to force the encoders to learn relevant sentence representations.

3   Model

*** TBC: Nein, was Du hier geschrieben hattest passt nicht: Hier gehoert die Kernidee von [1] hin; mein Ansatz war es eben genau diese Idee auf unser Problem anzuwenden (ein Satz -> ein Vektor). Hier wird nicht unterschieden zwischen den ersten 4 und dem 5. Satz; jeder Satz hat zwei Saetze als Kontext (ausser sinnigerweise der 1. und der 5. Satz einer story, die haben je nur einen). MaW, der ganze obere Teil von Seite 4 des papers bis und mit eqn (2) gehoert mit Zitierung hier hin. Deshalb passt auch die Grafik (a) weiter unten genau dazu und muss nicht angepasst werden. ***

For the encoding functions f and g, we settled on bidirectional RNNs with 600 GRU cells each.

When evaluating the trained model, the fourth story sentence is fed to the encoder f and the classifier then decides which of the two candidate endings encoded by encoder g is more probable (see figure).

Image of architecture:

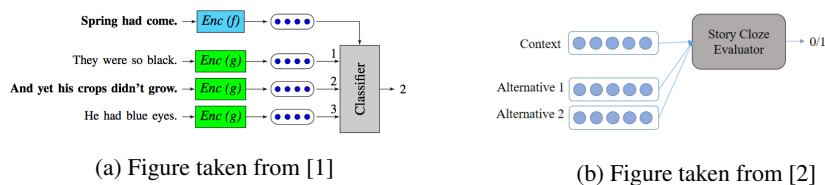(a) Figure taken from [1]   (b) Figure taken from [2]

Figure 1: **Frage:** Welche Art von Grafik ist für den Report geeigneter (links: müsste ich noch anpassen auf unsere Methode -NEIN, nicht anpassen, siehe oben; die Grafik rechts könnte man so übernehmen - diese Grafik sagt nichts relevantes ueber unseren Ansatz aus.

4 Training

For trainable word embeddings, we based on with GloVe pre-trained word embeddings in 300 dimensions [3]. As training data, we exclusively used the given corpus of 88161 five sentence stories. In these stories, there were no alternative (wrong) ending sentences. A context scope of +/- 1 was used, where the previous and the next sentences are predicted given an input sentence. After measuring accuracy on the validation set, we chose hyperparameters to be a minibatch size of 100 sentences (i.e. 20 cloze stories with 5 sentences each) with a dropout rate of 0.25. We trained over 5 epochs to minimize cross-entropy with Adam Optimizer at a learning rate of 1e-3. Due to the efficient nature of this approach, total training time on a single GTX 1080 Ti GPU was well under an hour.

5 Experiments

On the given validation set, we achieve an accuracy of 0.612 using the parameters mentioned above. This result is comparable with respect to what has been published earlier [].

6 Conclusion

Due to the limited scope and time of this project, we felt we couldn't fully exploit the benefits of our chosen approach. For further investigation, the use of more sophisticated embeddings (case senstive; Skip-thoughts) as well as using more sophisticated RNN encoding functions look promising.

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`.