

# CS288 HW4 Writeup

Will Lavanakul

## 1 Machine Translation

Two examples of where the model succeeds are listed below:

- Source: Jim und Carey: Das Geheimnis des Glücklichseins  
Target: Jim and Carey: The secret to being happy  
Model: Jim and Carey: The secret to happiness
- Source: Arthur Bailey: Kann Technologie all unsere Probleme lösen?  
Target: Arthur Bailey: Can technology solve all of our problems?  
Model: Arthur Bailey: Can technology solve our problems?

Two examples of where the model fails:

- Source: Danach gehen sie manchmal noch einkaufen  
Target: After that they sometimes go shopping.  
Model: Danelle: How to make your own food
- Source: Es freut mich, dich kennenzulernen  
Target: I am pleased to meet you.  
Model: David Deutsch: How to make free love

The main reason for this performance is due to the sources being in or out of distribution of the training data. Since we trained on the ted talk dataset, sources with the format ‘Author: Title’ are translated a lot better than general sentences. We can see in the failed examples, the model tries to translate to this format. Also, since the dataset is from Ted Talks, the subjects in the sources are much more diverse and technical, making it harder to generalize. The reason for the model outputting fluent sentences even when performing poorly is because the model was pretrained for language generation, then finetuned for translation. Even if the translation fails, the model still knows how to generate coherent english sentences.

Here are comparisons of the LSTM model and the finetuned GPT-2 model. The source is ‘Eine Gruppe von Männern lädt Baumwolle auf einen Lastwagen’ and the target is ‘A group of men are loading cotton onto a truck’.

- LSTM w/ Attn and greedy decoding: A group of men are loading off a truck.
- GPT-2 Finetuned: A group of young men in Germany build a last-mile bridge

On this example, the LSTM model performs better. Since the LSTM model was trained in the multi30k dataset, it performs better on a validation sample from the Multi30k dataset. The GPT-2 model is out of distribution.

## 2 Sentiment Analysis

There are several ways to improve this model. We can use data augmentation techniques (for example replacing words with synonyms) to generate more training examples. We can also use an ensemble of models. We can also use more datasets that are not just movie reviews. This would give the model a larger domain of data.