Swiss German Speech to Standard German Text

SwissText.org Shared Task 3

Alex Wolf
University of Zurich
alex.wolf@uzh.ch

Deborah Noemie Jakobi
University of Zurich
deborahnoemie.jakobi@uzh.ch

Abstract

This paper analyzes and implemented models for the shared task 3 from the SwissText conference on translating swiss german speech to standard german text and presents the findings. We implemented multiple *DeepSpeech* models using the data provided by SwissText.org as well as additional data. An additional experiment with a sequence to sequence translation model was trained in order to improve our score. We achieved a BLEU score of up to 0.17 on the test set of SwissText.

1 Introduction

Swiss German has a wide variety of different dialects, with a huge difference in words, pronunciation, even to the point of sounding like a different language. Tackling a standardized translation of different spoken Swiss German dialects into standardized German text requires a vast amount of data and fine tuning. The SwissText Conference (STC) 2021 proposed a shared task to tackle this problem and provided a dataset (SwissText Conference dataset (STCD) to train and fine tune on. This paper shows what kind of experiments, data, and approaches the authors used to tackle this problem.

2 Literature review

The shared task of a translating spoken Swiss German into standard written German was already presented by the STC in 2020. While the current 2021 task is about reaching the highest BLEU score, last year's submissions where ranked based on the least Word Error Rate (WER). **?** achieved the best WER of 40.29%. The authors used a CNN acoustic model named Jasper. They used additional Standard German data and fine-tuned on the official data set. They used different augmentation techniques and a language model. **?** add text. (**?**) achieved the second best WER. ... **?** used an end-to-end model called DeepSpeech and achieved an WER of 58.93%. This is the model we decided to use as well. It is publicly available on GitHub and can easily be fine-tuned or used for transfer learning.

3 DeepSpeech

explain how deepspeech works...

4 Experiments

4.1 Datasets

We compared models on both the dataset provided by the STC as well as the ArchiMob corpus. The STCD contains 38 GB of labeled and 65 GB of unlabeled spoken swiss german audio data and an additional validation set containing 1.5 GB of data (**?**). The ArchiMob corpus (Release 2) contains X GB of spoken swiss german data (**?**).

4.2 Experiments

The base model of most our experiments was a pretrained DeepSpeech model (release 9.0, get link!). The first model we trained served as our private baseline. We trained a bare DeepSpeech model on the SwissText labelled dataset.

5 Results & Discussion

Interestingly, we noticed that our internal deviations in the BLEU score are not as great as on the official test set.

Model#	Data	Train BLEU	Test BLEU
1	SwissText	0.23	0.0004
2	ArchiMob	0.27	**0.17**
3	ArchiMob	0.24	0.07

Table 1: Font guide.

6 Conclusion