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) 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.
(Pluss et al., 2020) add text.
(Buchi et al., 2020) add text.
(Kew et al., 2020) add text.

3 Materials and Methods

4 Experiments & Results

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 (Ando and Zhang, 2005). The ArchiMob corpus (Release 2) contains X GB of spoken swiss german data (Samardžić et al., 2016).

4.2 Results

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.

5 Discussion

6 Conclusion

References

Rie Kubota Ando and Tong Zhang. 2005. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853.

Matthias Buchi, Malgorzata Anna Ulasik, Manuela Hürlimann, Fernando Benites, Pius von Daniken, and Mark Cieliebak. 2020. Zhaw-init at germeval 2020 task 4: Low-resource speech-to-text.

Tannon Kew, Iuliia Nigmatulina, Lorenz Nagele, and Tanja Samardžić. 2020. Uzh tilt: A kaldi recipe for swiss german speech to standard german text.

Michel Pluss, Lukas Neukom, and Manfred Vogel. 2020. Germeval 2020 task 4: Low-resource speech-to-text.

Tanja Samardžić, Yves Scherrer, and Elvira Glaser. 2016. ArchiMob - a corpus of spoken Swiss German. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4061–4066, Portorož, Slovenia. European Language Resources Association (ELRA).