Linguistically Motivated Sign Language Segmentation

Amit Moryossef, Zifan Jiang, Mathias Müller, Sarah Ebling, Yoav Goldberg

Sign language segmentation is a crucial task in sign language processing systems. It enables downstream tasks such as sign recognition, transcription, and machine translation. In this work, we consider two kinds of segmentation: segmentation into individual signs and segmentation into *phrases*, larger units comprising several signs. We propose a novel approach to jointly model these two tasks. Our method is motivated by linguistic cues observed in sign language corpora. We replace the predominant IO tagging scheme with BIO tagging to account for continuous signing. Given that prosody plays a significant role in phrase boundaries, we explore the use of optical flow features. We also provide an extensive analysis of hand shapes and 3D hand normalization. We find that introducing BIO tagging is necessary to model sign boundaries. Explicitly encoding prosody by optical flow improves segmentation in shallow models, but its contribution is negligible in deeper models. Careful tuning of the decoding algorithm atop the models further improves the segmentation quality. We demonstrate that our final models generalize to out-of-domain video content in a different signed language, even under a zero-shot setting. We observe that including optical flow and 3D hand normalization enhances the robustness of the model in this context.