Convolutional network: baseline of state-of-the-art quality for low-resource morpheme segmentation.

Anonymous ACL submission

Abstract

We apply convolutional neural networks to the task of shallow morpheme segmentation using low-resource datasets for 5 different languages. We show that both in fully supervised and semi-supervised settings our model beats previous state-of-the-art approaches. We argue that convolutional neural networks reflect local nature of morpheme segmentation better than other semi-supervised approaches.

Morpheme segmentation consists in dividing a given word to meaningful individual units, morphemes. For example, a word *unexpectedly* could be segmented as *un-expect-ed-ly*. The generated segmentation may be used as input representation for machine translation (**?**) or morphological tagging (**?**) or for automatic annotation of digital linguistic resources.