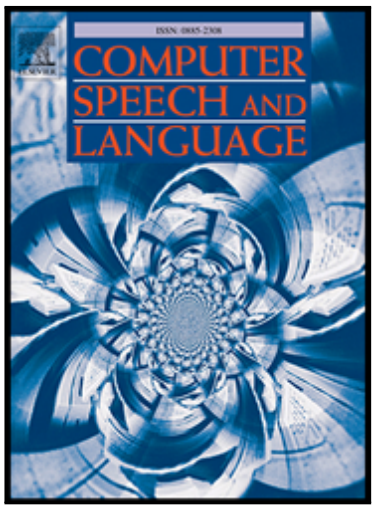

On Integrating a Language Model Into Neural Machine Translation

Caglar Gulcehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho,
Yoshua Bengio

PII: S0885-2308(16)30139-5
DOI: 10.1016/j.csl.2017.01.014
Reference: YCSLA 836

To appear in: *Computer Speech & Language*

Received date: 1 May 2016
Revised date: 21 November 2016
Accepted date: 10 January 2017

Please cite this article as: Caglar Gulcehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, Yoshua Bengio, On Integrating a Language Model Into Neural Machine Translation, *Computer Speech & Language* (2017), doi: 10.1016/j.csl.2017.01.014

On Integrating a Language Model Into Neural Machine Translation

Caglar Gulcehre[a,1,*], Orhan Firat[b,1], Kelvin Xu[a], Kyunghyun Cho[c,a], Yoshua Bengio[a,d]

[a] *University of Montreal, Canada*
[b] *Middle East Technical University, Turkey*
[c] *New York University, USA*
[d] *CIFAR Senior Fellow*

Abstract

Recent advances in end-to-end neural machine translation models have achieved promising results on high-resource language pairs such as En→Fr and En→De. One of the major factor behind these successes is the availability of high quality parallel corpora. We explore two strategies on leveraging abundant amount of monolingual data for neural machine translation. We observe improvements by both combining scores from neural language model trained only on target monolingual data with neural machine translation model and fusing hidden-states of these two models. We obtain up to 2 BLEU improvement over hierarchical and phrase-based baseline on low-resource language pair, Turkish→English. Our method was initially motivated towards tasks with less parallel data, but we also show that it extends to high resource languages such as Cs→En and De→En translation tasks, where we obtain 0.39 and 0.47 BLEU improvements over the neural

☆This paper is an extended version of an earlier technical report [1].

*Corresponding author
Email address: caglar.gulcehre@umontreal.ca (Caglar Gulcehre)

[1]Equally contributed