arXiv:2203.01927 (cs)

[Submitted on 3 Mar 2022]

As Little as Possible, as Much as Necessary: Detecting Over- and Undertranslations with Contrastive Conditioning

Jannis Vamvas, Rico Sennrich

Omission and addition of content is a typical issue in neural machine translation. We propose a method for detecting such phenomena with off-the-shelf translation models. Using contrastive conditioning, we compare the likelihood of a full sequence under a translation model to the likelihood of its parts, given the corresponding source or target sequence. This allows to pinpoint superfluous words in the translation and untranslated words in the source even in the absence of a reference translation. The accuracy of our method is comparable to a supervised method that requires a custom quality estimation model.

Comments: ACL 2022
Subjects: **Computation and Language (cs.CL)**
Cite as: arXiv:2203.01927 **[cs.CL]**
(or arXiv:2203.01927v1 **[cs.CL]** for this version)
https://doi.org/10.48550/arXiv.2203.01927

Submission history

From: Jannis Vamvas [view email]
[v1] Thu, 3 Mar 2022 18:59:02 UTC (98 KB)

Code, Data, Media

Demos

Related Papers

About arXivLabs

Which authors of this paper are endorsers? | *Disable MathJax* (*What is MathJax?*)