Abstract

Massive open online courses have been growing rapidly in size and impact. TraMOOC\(^1\) aims at developing high-quality translation of all types of text genre included in MOOCs from English into eleven European and BRIC languages that are hard to translate into and have weak MT support.

1 Recent developments

In TraMOOC, we have developed machine translation prototypes for 11 target languages, from English into German, Italian, Portuguese, Dutch, Bulgarian, Greek, Polish, Czech, Croatian, Russian, and Chinese. The translation systems are based on phrase-based SMT and neural machine translation. The latter has achieved state-of-the-art performance in recent evaluation campaigns (Bojar, 2016). We use the Nematus toolkit (Sennrich, 2017) for training; the translation server is based on the amuNMT toolkit (Junczys-Dowmunt et al., 2016). The translation systems have been adapted to MOOC texts via fine-tuning of the model parameters on in-domain training data to maximize translation quality on this domain.

\(^1\)TraMOOC is a H2020 Innovation Action project funded by the European Commission (H2020-ICT-2014-1-ICT-2014/644333) and runs from February 2015 to February 2018. For more details on the project, please, visit http://www.tramooc.eu

References

