

TeacherSim: Cross-lingual Machine Translation Evaluation with Monolingual Embedding as Teacher

Hao Yang^a, Min Zhang^a, Shimin Tao^a, Miaomiao Ma^a, Yin Qin^a, Daimeng Wei^a

^a2012 Labs, Huawei Technologies CO., LTD

Corresponding Author: yanghao30@huawei.com, Tel: +86 135 2048 8048

Keywords: Machine Translation Evaluation, Knowledge Distillation, Monolingual Embedding

Abstract:

Cross-lingual machine translation (MT) evaluation or quality estimation is usually performed either from cross-lingual transfer in supervised tasks or via unsupervised multilingual textual similarity. XMoverScore investigates a range of state-of-the-art cross-lingual evaluations by feature extraction on token-level similarity and multilingual pre-trained models mBERT and LASER. In this paper, monolingual and sentence-level embedding is used as a teacher encoder (TeacherSim) to align cross-lingual MT sentences. It is found that the monolingual teacher encoder can enhance both quality and usability in that: (a) A monolingual embedding teacher assures cross-lingual semantic alignment and enhances evaluation performance. (b) The last layer always delivers the best performance based on the Sentence Transformer architecture, while layer selection is sensitive for different models with BERTScore. Experiments like segment-level MT evaluation show that TeacherSim surpasses most strong baselines for all language pairs..



Hao Yang, Research Scientist of 2012 Lab, Huawei Company limited. This author became a Member (M) of IEEE in 2005, a Senior Member (SM) in 2009. Dr Yang achieved Phd from Beijing University of Posts and Telecommunications in 2009, and majoring research fields includes Natural Language Processing, Neural Network Translation and Deep Learning in Text areas.