Improved Statistical Machine Translation for Resource-Poor Languages Using Related Resource-Rich Languages

We propose a novel language-independent approach for improving statistical machine translation for resource-poor languages by exploiting their similarity to resource-rich ones. More precisely, we improve the translation from a resource-poor source language X1 into a resource-rich language Y given a bi-text containing a limited number of parallel sentences for X1-Y and a larger bi-text for X2-Y for some resource-rich language X2 that is closely related to X1. The evaluation for Indonesian->English (using Malay) and Spanish->English (using Portuguese and pretending Spanish is resource-poor) shows an absolute gain of up to 1.35 and 3.37 Bleu points, respectively, which is an improvement over the rivaling approaches while using much less additional data.