Analysis of Experimentation

[Saahas Parise]

[March 20, 2024]

In my experimentation, I modified the training data using back translation (BT). I created a model using attention that translated English text to German; I then augmented the existing German, English training data with the German, English data from the new model. I also experimented with implementing weight decay (WD) of 1e-5. Because the training dataset is only about 29000 examples, I felt that augmenting this data would improve model training and generalization, therefore raising BLEU. Also, the motivation behind using weight decay was to help the baseline (B) and attention (A) models generalize better to new data (such as the validation and testing sets) through reducing overfitting. From the table below, we can observe that for both attention and baseline experiments, weight decay and back translation used in tandem reached higher BLEU scores than those without weight decay or back translation. This shows that the combination of the added features improved machine translation performance on the validation set. However, weight decay seemed to have a much greater effect on BLEU than back translation. Overall, the attention model with weight decay and back translation reached a higher BLEU score than the original model ("A w/out weight decay, w/out BT") by ~ 1.068, meaning the extension was successful.

Model	BLEU Score
B w/out WD, w/out BT	21.215
B w/ WD, w/out BT	24.488
B w/out WD, w/ BT	21.932
B w/ WD, w/ BT	24.687
A w/out WD, w/out BT	38.898
A w/ WD, w/out BT	39.111
A w/out WD, w/ BT	38.308
A w/ WD, w/ BT	39.966

In this table, B means baseline model, A means attention model, WD refers to adding weight decay, and BT refers to back translation.