

NLP Applications

Project : Neural System Combination for Machine Translation

Mid Report

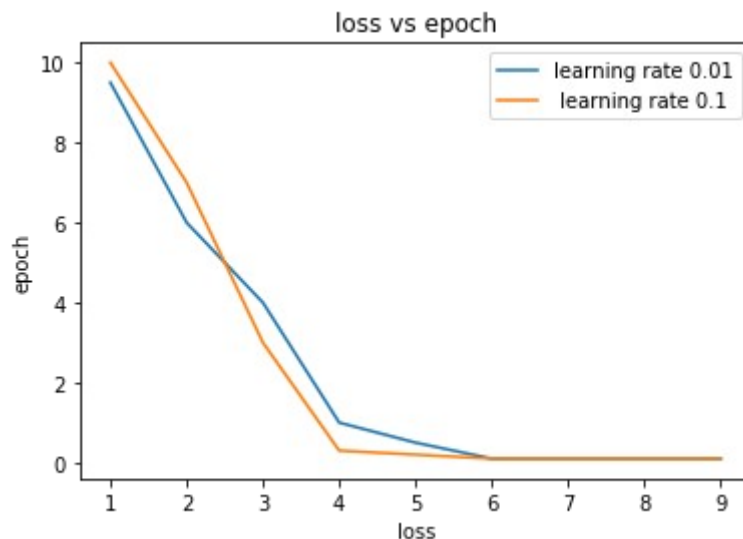
Submitted by:
Chittaranjan Rath 2018201007
Suraj Garg 2018202003

Introduction

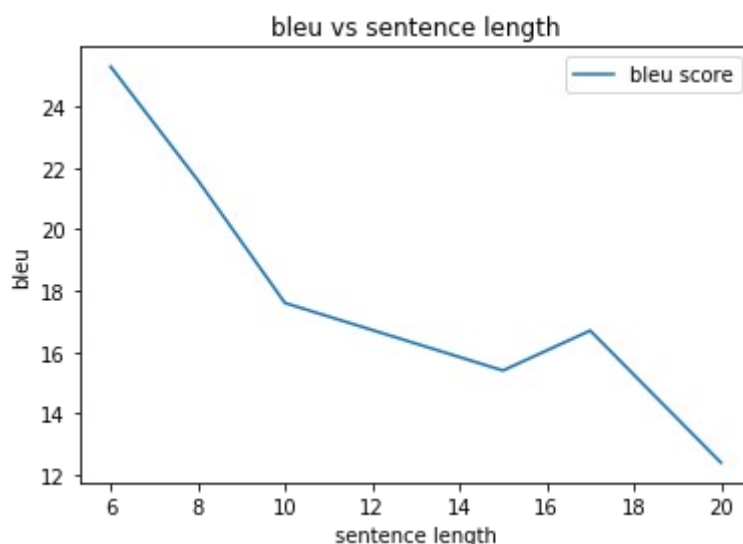
In this project we needed to run attention based model on the dataset and also run phrase based model using mooses on the dataset . Then we need to combine the translations so as to get the advantages of both NMT and SMT. To combine we use neural system combination. The dataset was divided into half for training and testing purposes.

Attention based model:

loss vs epoch



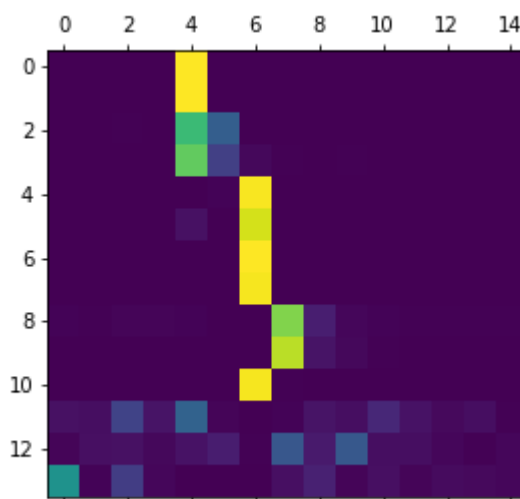
Bleu score vs length of sentences



Attention Weight visualization:

['Every', 'thing', 'about', 'Maldiv', 'was', 'overwhelming', '.']

मालदीव के बारे में हर बात अभित करने वाली थी ।



Example tranlations from test set:

Highlight border color

सीमांत (बोर्डर) के रंग को हाइलाइट करें

Script Recorder

लिपि रेकोर्डर

Tests fundamental GUI application accessibility

मूलभूत जीयूआई अनुप्रयोग पहुंचनीयता का परीक्षण करता है

The duration of the highlight box when selecting accessible nodes

पहुंचनीय आसंधि (नोड) को चुनते समय हाइलाइट बक्से की अवधि

For Phrase based model Moses was used

We installed moses and used GIZA++ 16 , for word-aligning our parallel corpus
ans KenLM 19 for language model estimation.

To prepare the data for training the translation system, we have performed the following steps:

- tokenisation: This means that spaces have to be inserted between (e.g.) words and punctuation.
 - truecasing: The initial words in each sentence are converted to their most probable casing. This helps reduce data sparsity.
 - cleaning: Long sentences and empty sentences are removed as they can cause problems with the training pipeline, and obviously mis-aligned sentences are removed.
- 3-gram language model was built using KenLM.

Example tranlations from test set:

The default plugin layout for the top panel

ऊपरी पटल के लिए डिफोल्ट प्लग-इन खाका

Highlight duration

अवधि को हाइलाइट रकें

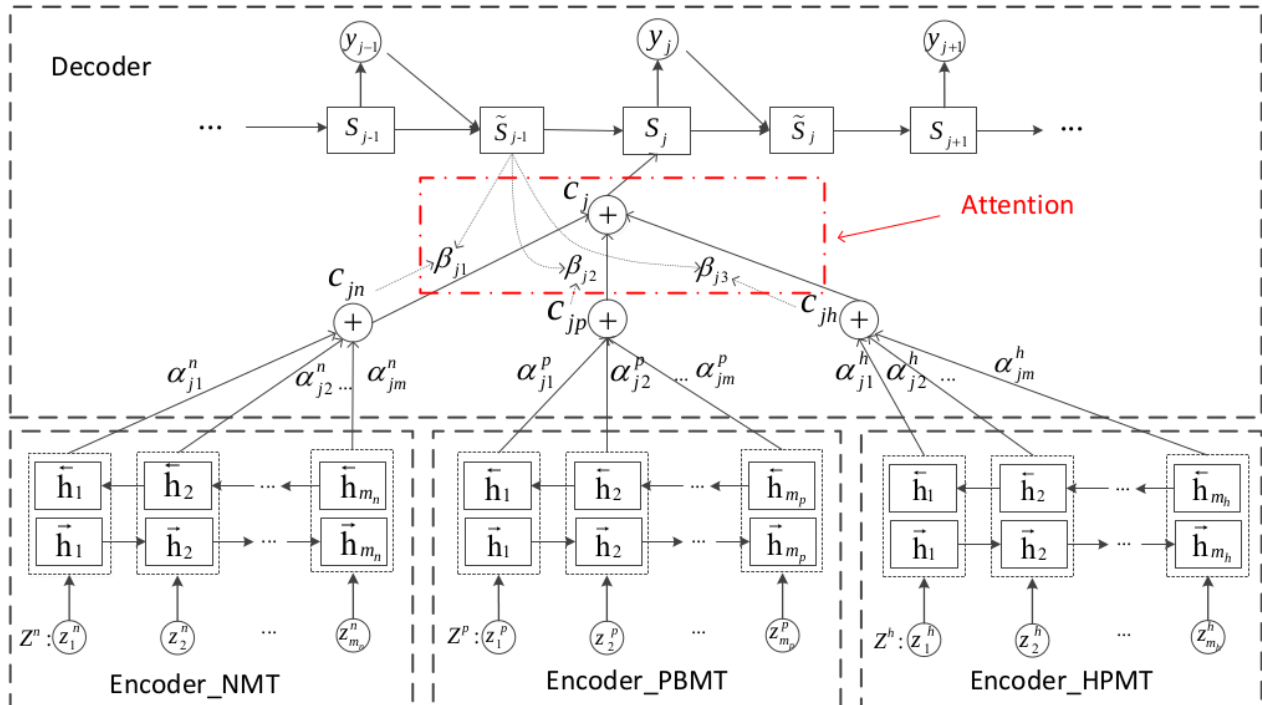
The color and opacity of the highlight fill.

हाइलाइट किया गया भराई का रंग और पारदर्शिता।

Highlight fill color
भराई के रंग को हाइलाइट करें

To Do:

Combine the outputs of both the above models and prepare a new parallel corpus and apply attention over it to combine them according to following figure. Then prepare complete report and show results



Github: <https://github.com/chittaranjan-rath/Neural-System-Combination-for-Machine-Translation>