Grammatical Error Correction using Neural Networks

Yokila Arora, Ayush Gupta, Jaspreet Kaur

{yarora, ayushg, j031723}@stanford.edu

Objective

- Mitigating the limitations of providing manual writing feedback for English language learners by automation of Grammatical Error Correction (GEC)
- While rule-based models and SMT have been the popular choice, neural networks have the potential to improve results by leveraging global context and capture non-linear relationships
- The presented work integrates Statistical Machine Translation (SMT) model with a Neural Network Global Lexicon Model (NNGLM)
- The proposed model enhances SMT baseline performance by $F_{0.5}$ score of 0.57

Dataset

- **Training**: NUCLE (NUS Corpus of Learner English)
- **Development:** Test set for CONLL 2013 shared task
- **Test:** Test set for CONLL 2014 shared task

	Train data	Dev. data	Test data
# essays	1,397	50	50
# sentences	57,151	1,381	1,312
# word tokens	1,61,567	29,207	30,144

- **SMT Translation Model:**
  NUCLE, Lang-8 Corpus of Learner English v1.0
- **SMT Language Model:**
  English Wikipedia (~1.78 billion tokens)

Preprocessing

- Source and Target vocabularies were generated for a binary bag of words representation after information extraction from the Corpus
- Sentence pairs which were incomplete, or too long (> 80 tokens), or offered no corrections, were removed

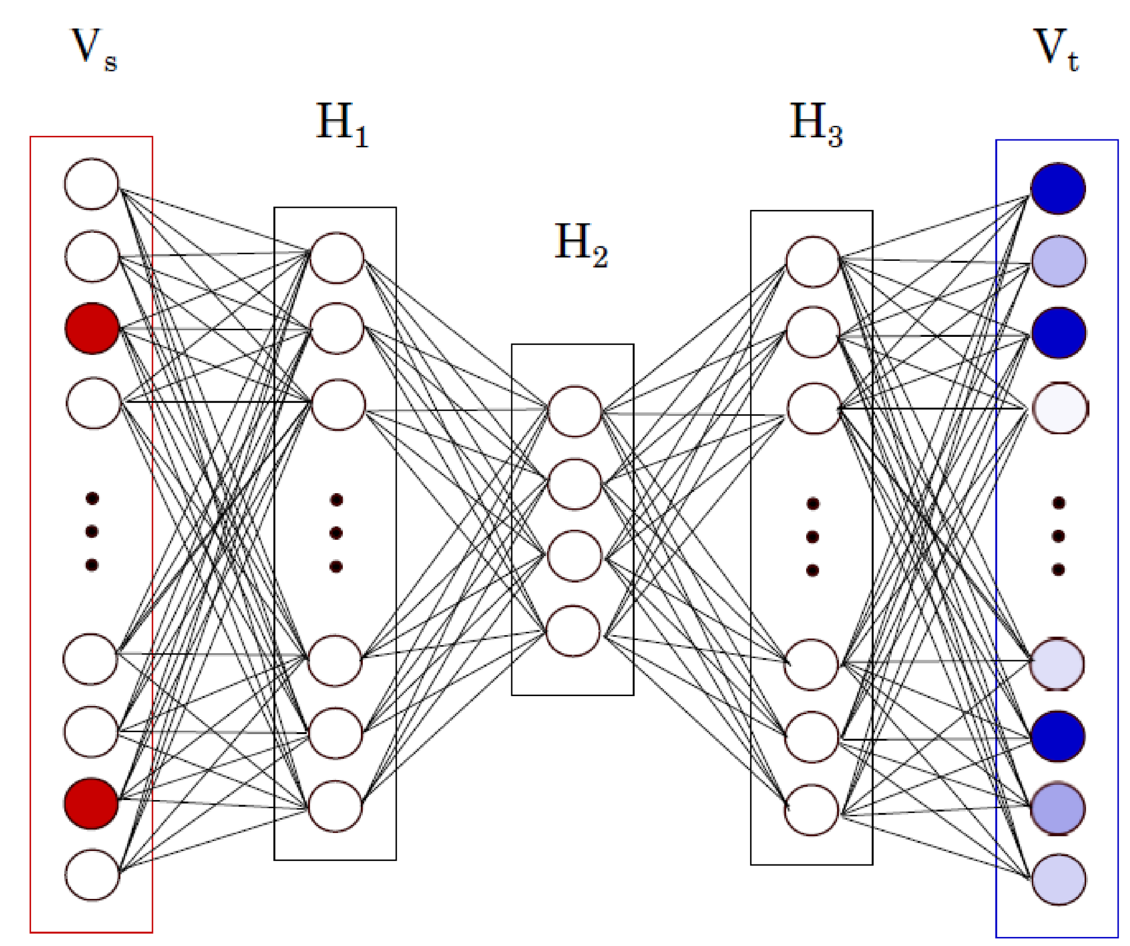
Model

- Target hypothesis were generated by SMT using default features in Moses, further scored by NNGLM which predicts the presence of words in the corrected output
- NNGLM estimates the overall probability of a target hypothesis given the source sentence by concatenating the produced probabilities of individual words present in the sentence, as:

$$P(T_h|S) \approx \prod_{j=1}^{|T|} p(t_j|S) \quad \forall h$$

- The individual word probabilities $P(t|S)$ are outputs of the NNGLM which is a feed forward neural network with 3 hidden layers trained using mini-batch gradient descent to minimize cross entropy loss.
- NNGLM architecture was chosen after investigating impact of network configurations like no. of hidden layers, no. of neurons, learning rate, and batch size within computational limitations

NNGLM Architecture



$|H_1| = 1000;\ |H_2| = 500;\ |H_3| = 1000;\ \alpha = 0.02;$ #Batch: 15

$$p(t|S) = \sigma\left(W^{[4]}.a^{[3]} + b^{[4]}\right)$$

$$E = \frac{-1}{|V_T|}\sum_{j=1}^{|V_T|}\left[\hat{T}_j \log p(t_j|S) + (1-\hat{T}_j)\log\left(1 - p(t_j|S)\right)\right]$$

Results

Results on training, development, and test data

	Training	Dev.	Test
P	52.26	50.84	50.76
R	23.67	23.39	23.43
$F_{0.5}$	42.09	41.17	41.15

Comparison of results against SMT Baseline, and Ref. [1]

	SMT baseline	Ref. [1]	Proposed
P	50.56	22.68	50.76
R	22.68	23.21	23.43
$F_{0.5}$	40.58	41.01	41.15

Sample system output for an input from test data

- **Input:** Above all, life is more important than *secret*.
- **Output:** Above all, life is more important than secrets.
- **Reference:** Above all, life is more important than secrets.

Discussion

- The ability of NNGLM to model words and phrases in continuous phase, provide global context and capture non-linear mapping, enables them to generalize better and provide improved GEC corrections.

Future Work

- Integrating Neural Network Joint Model on the SMT baseline along with NNGLM can offer further improvement
- The $F_{0.5}$ score of the proposed model can be enhanced with use of more NLP tools and extensive training on varied data

References

1. Chollampatt, Shamil, Kaveh Taghipour, and Hwee Tou Ng. "Neural network translation models for grammatical error correction." *arXiv preprint arXiv:1606.00189* (2016).
2. Ha, Thanh-Le, Jan Niehues, and Alex Waibel. "Lexical translation model using a deep neural network architecture." *arXiv preprint arXiv:1504.07395* (2015).