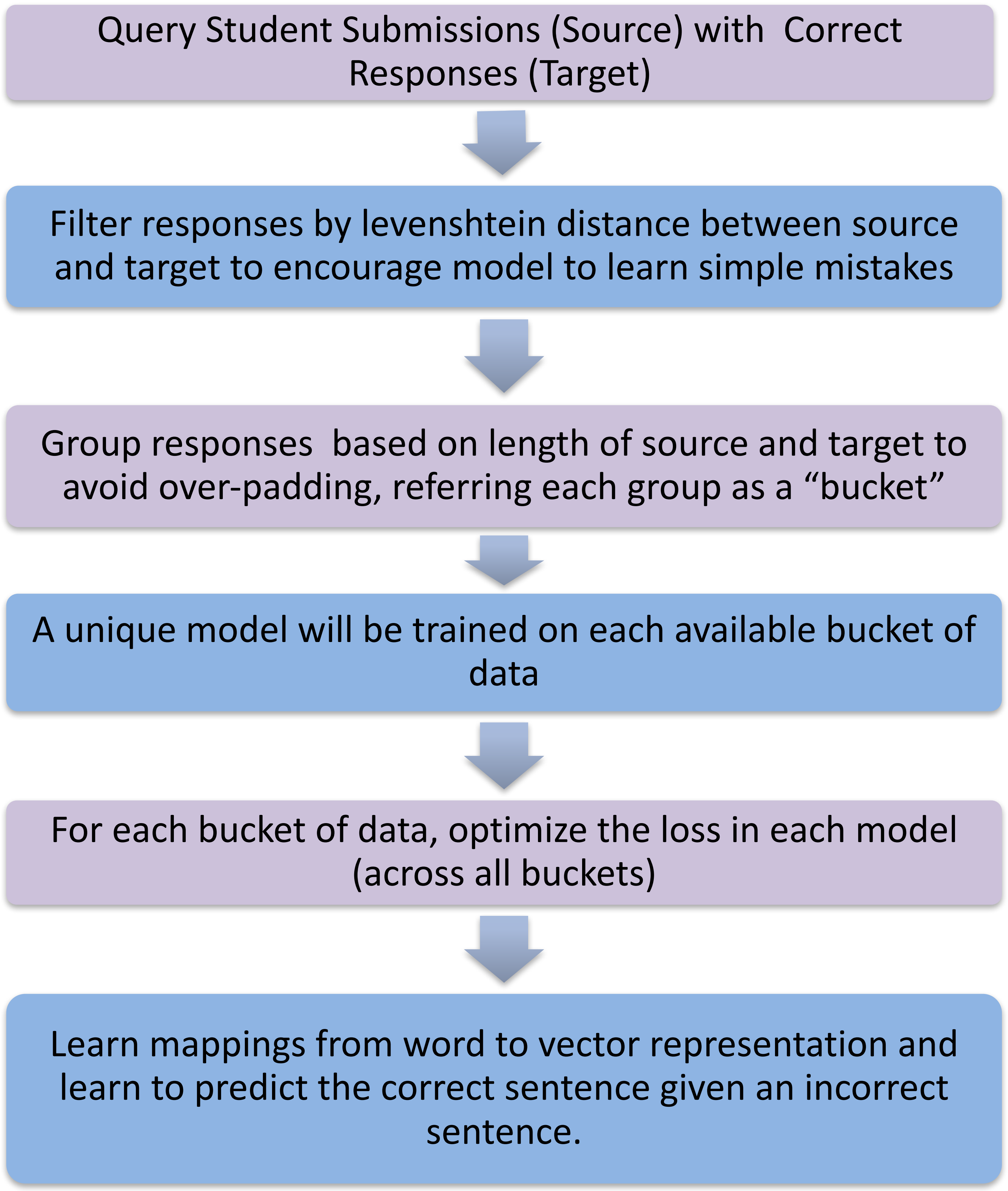
Seq2Seq Modeling for Grammar Correction

Emanuel Cortes
Department of Computer Science, Stanford University

Motivation

- Develop model to correct simple grammatical errors in sentences.
- Goal: To create a writing tool to help anyone learn grammar and improve their grammar skills when working on a writing activity .

Data Preparation and Training Pipeline



Query Student Submissions (Source) with Correct Responses (Target)

Filter responses by levenshtein distance between source and target to encourage model to learn simple mistakes

Group responses based on length of source and target to avoid over-padding, referring each group as a "bucket"

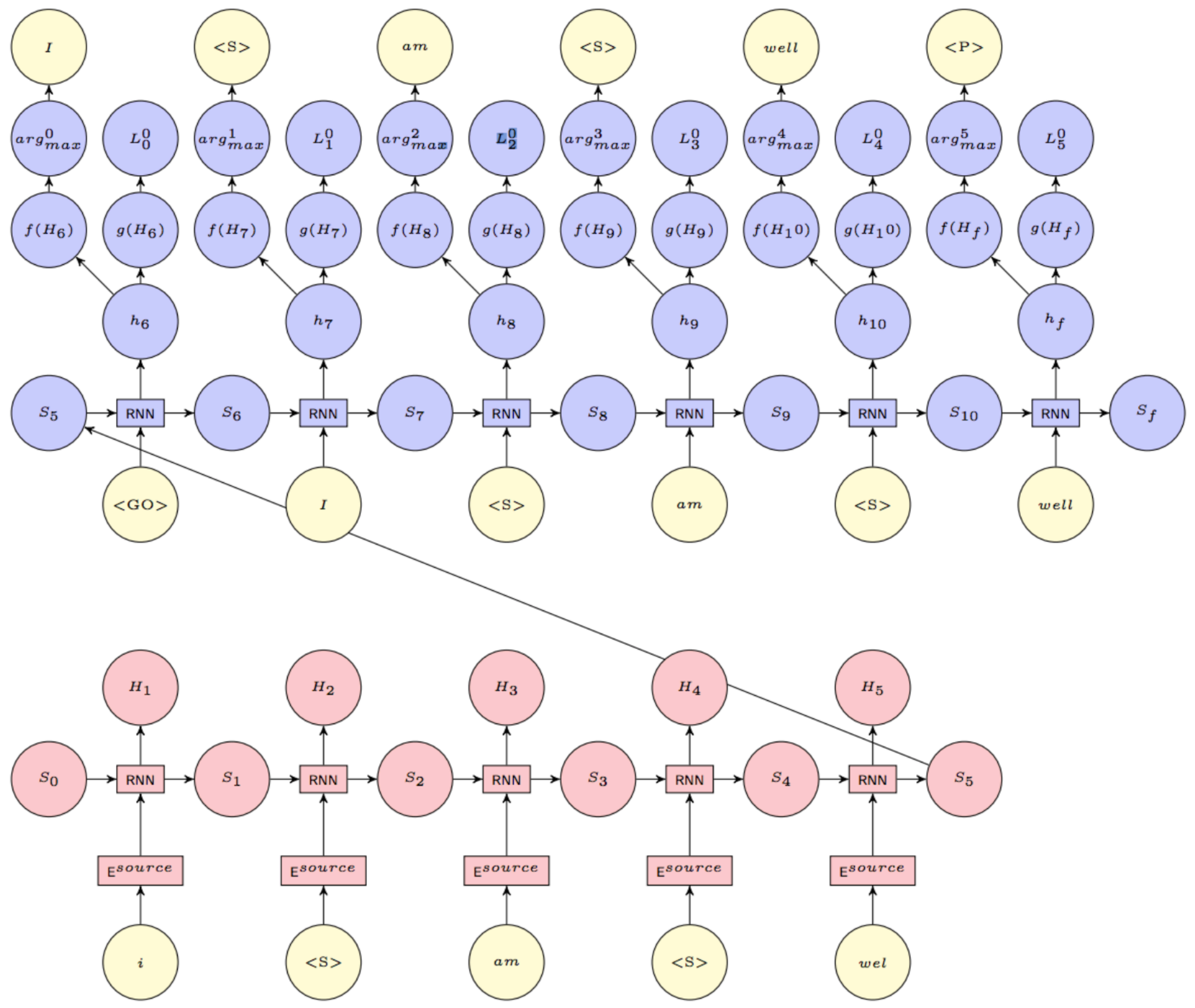
A unique model will be trained on each available bucket of data

For each bucket of data, optimize the loss in each model (across all buckets)

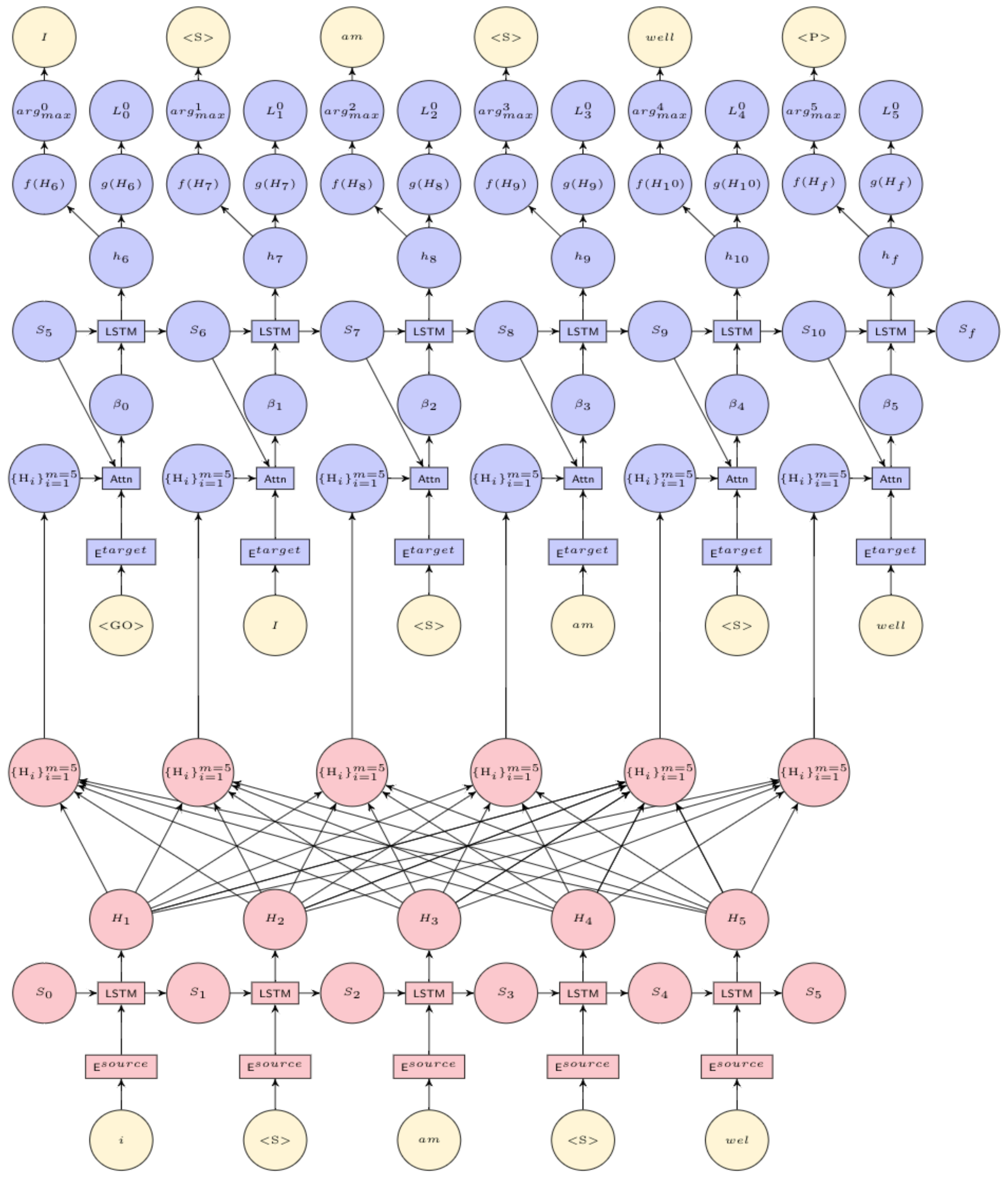
Learn mappings from word to vector representation and learn to predict the correct sentence given an incorrect sentence.

Models

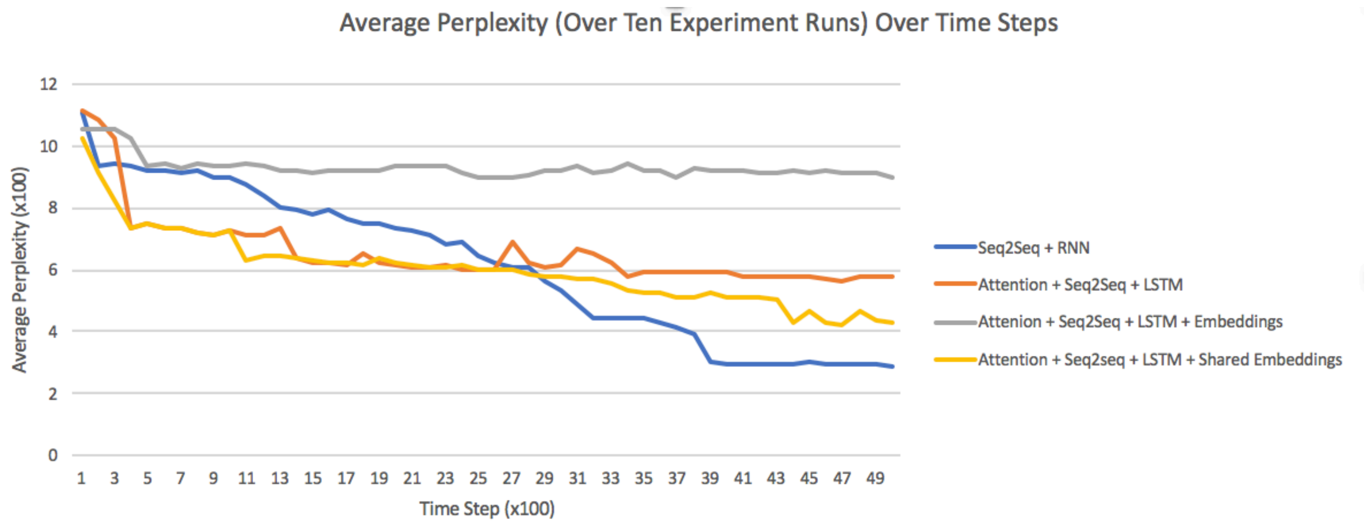
1.) Baseline:


I
<S>
am
<S>
well
<P>
arg^0_{max} L^0_0 arg^1_{max} L^0_1 arg^2_{max} L^0_2 arg^3_{max} L^0_3 arg^4_{max} L^0_4 arg^5_{max} L^0_5
$f(H_6)$ $g(H_6)$ $f(H_7)$ $g(H_7)$ $f(H_8)$ $g(H_8)$ $f(H_9)$ $g(H_9)$ $f(H_10)$ $g(H_10)$ $f(H_f)$ $g(H_f)$
h_6 h_7 h_8 h_9 h_{10} h_f
S_5 RNN S_6 RNN S_7 RNN S_8 RNN S_9 RNN S_{10} RNN S_f
<GO> I <S> am <S> well
H_1 H_2 H_3 H_4 H_5
S_0 RNN S_1 RNN S_2 RNN S_3 RNN S_4 RNN S_5
E^{source}
i
<S>
am
<S>
wel

2.) Attention Seq2Seq:


I
<S>
am
<S>
well
<P>
arg^0_{max} L^0_0 arg^1_{max} L^0_1 arg^2_{max} L^0_2 arg^3_{max} L^0_3 arg^4_{max} L^0_4 arg^5_{max} L^0_5
$f(H_6)$ $g(H_6)$ $f(H_7)$ $g(H_7)$ $f(H_8)$ $g(H_8)$ $f(H_9)$ $g(H_9)$ $f(H_{10})$ $g(H_{10})$ $f(H_f)$ $g(H_f)$
h_6 h_7 h_8 h_9 h_{10} h_f
S_5 LSTM S_6 LSTM S_7 LSTM S_8 LSTM S_9 LSTM S_{10} LSTM S_f
a_0 a_1 a_2 a_3 a_4 a_5
$(H_i)_{i=1}^{m=5}$ Attn
E^{target}
<GO> I <S> am <S> well
$(H_i)_{i=1}^{m=5}$
H_1 H_2 H_3 H_4 H_5
S_0 LSTM S_1 LSTM S_2 LSTM S_3 LSTM S_4 LSTM S_5
E^{source}
i
<S>
am
<S>
wel

Experimental Results



- All average BLEU scores are near zero.
- For attention seq2seq + LSTM + shared embeddings, model is able to correct capitalization mistakes for smaller sentences.

Discussion

- Attention seq2seq model with embeddings may need to be trained for a longer period of time to achieve better performance.
- The model seems to train faster when sharing embeddings between encoder and decoder.
- With limited gpu/cpu resources, I need to balance expressive power of model with time complexity.

Future Work

- Less Complexity: Implement Sampled Softmax and Share Variables Across All Buckets.