

Sequence to Sequence Learning with Neural Networks - Cons

Suskever, Vinyals, Le

Presentation by Ryan Wolter

Goals - What is your main contribution? Sell that.

- “general end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure.”
 - Not so general because you only did two tasks, both of which were language related.
 - Speech to Text, audio processing, etc

Doesn't Age Well

- Main Contributions
 - Deep LSTM for seq2seq
 - SMT Hypotheses Reordering
 - Sentence reversing
- THIS LIVES IN A VOID

lation task from the WMT11-12 test set. The model achieved a BLEU score of 34.8 on the test set, which was penalized on out of vocabulary words.

On long sentences, the model achieved a BLEU score of 33.3 on the 1000 hypotheses produced.

to rerank the 1000 hypotheses produced. The BLEU score increases to 36.5, which shows that the LSTM also learned sensible phrases.

Not State of the Art

and is close to the previous state-of-the-art (which is 37.0 [9]).

But what is?

pairs. We have improved upon our 2013 system by i) using generalized representations, specifically automatic word clusters for translations out of English, ii) using unsupervised character-based models to translate unknown words in Russian-English and Hindi-English pairs, iii) synthesizing Hindi data from closely-related Urdu data, and iv) building huge language models on the common crawl corpus.

This is actually pretty complicated, sell THIS.

Main Contributions

Deep LSTM for seq2seq

Is there a shallow LSTM to compare to?

Method	test BLEU score (ntst14)
Bahdanau et al. [2]	28.45
Baseline System [29]	33.30
Single forward LSTM, beam size 12	26.17
Single reversed LSTM, beam size 12	30.59
Ensemble of 5 reversed LSTMs, beam size 1	33.00
Ensemble of 2 reversed LSTMs, beam size 12	33.27
Ensemble of 5 reversed LSTMs, beam size 2	34.50
Ensemble of 5 reversed LSTMs, beam size 12	34.81

Maybe your main contribution is the value of creating ensembles of 5 LSTMs...

Main Contributions

- Sentence Reversal
 - Kinda sounds like a hack, so maybe work harder to explain its importance?
 - Attention may solve this?
 - Although, the authors did say that they didn't have issues with long sentences
 - Sentence Reversal is NOT general end-to-end seq2seq...

Null Pointer Exception



and is close to the previous state-of-the-art (which is 37.0 [9]).

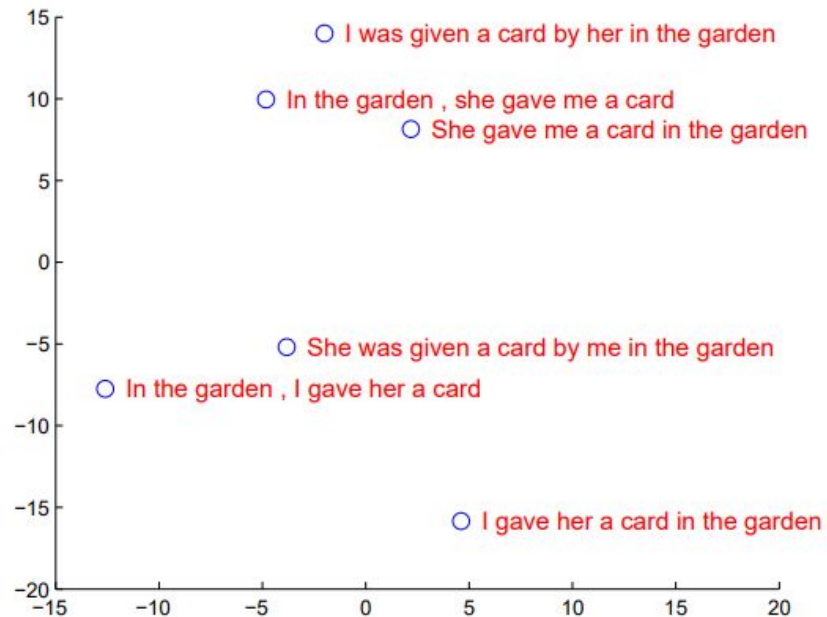
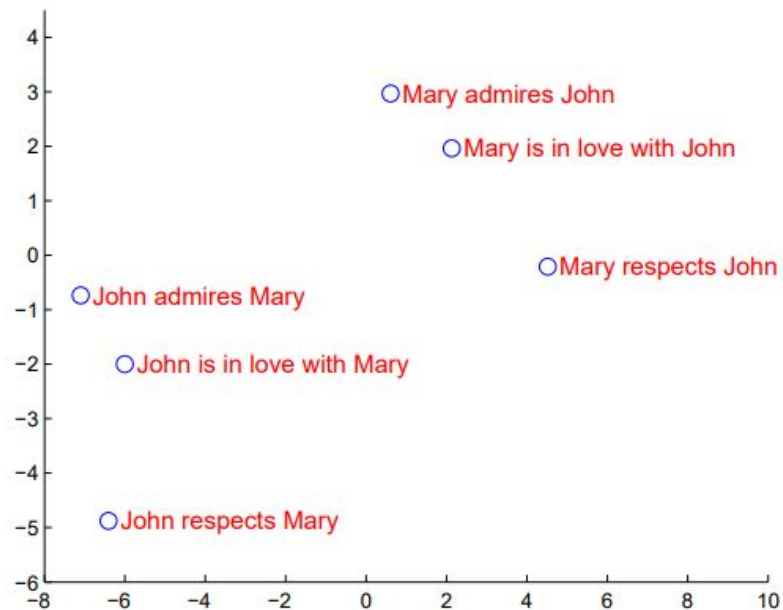
of a SMT baseline on this dataset is 33.30 [29].
SMT with a vocabulary of 80k words so the score

Method	test BLEU score (ntst14)
<u>Baseline System</u> [29]	33.30
<u>Cho et al.</u> [5]	34.54
<u>State of the art</u> [9]	37.0

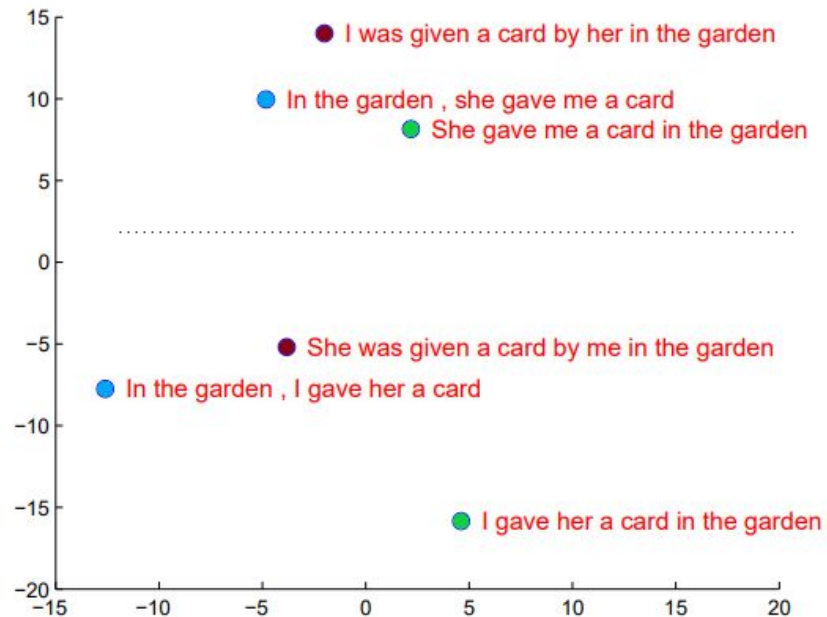
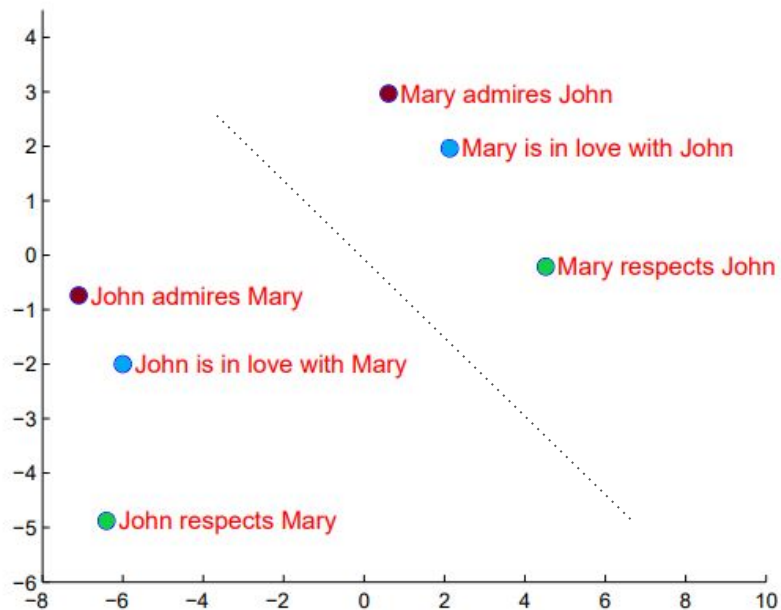
2-dimensional PCA projection of
the figures. The phrases are cluster

This is not as bad..
But still

Results Presentation



Results Presentation



Results Presentation

Our model	Avec la crémation , il y a un “ sentiment de violence contre le corps d’ un être cher ” , qui sera “ réduit à une pile de cendres ” en très peu de temps au lieu d’ un processus de décomposition “ qui accompagnera les étapes du deuil ” .
Truth	Il y a , avec la crémation , “ une violence faite au corps aimé ” , qui va être “ réduit à un tas de cendres ” en très peu de temps , et non après un processus de décomposition , qui “ accompagnerait les phases du deuil ” .

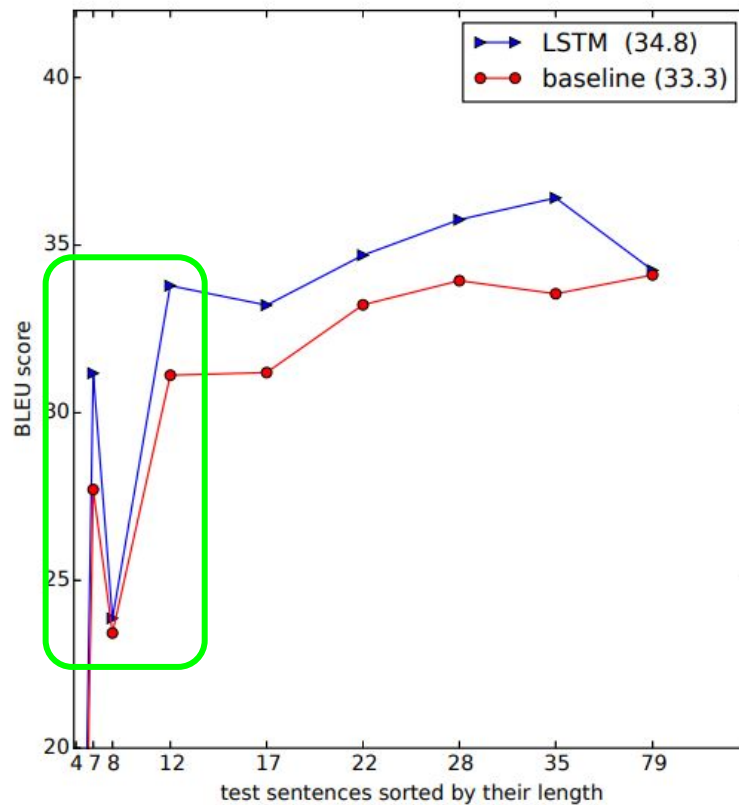
Table 3: A few examples of long translations produced by the LSTM alongside the ground truth translations. The reader can verify that the translations are sensible using Google translate.

Suggestion: Have a Human do back translation or give some kind of result on this, rather expecting us to google translate...

Results Presentation

What happened here?

(I'm just curious)



GPU Optimization

- “This was too slow for our purposes”
 - What are your purposes?
 - 8x GPU increase led to a 4x speed increase, 4x is not a huge order of magnitude...
- 4 layers by 1000 cells with 1000 word embeddings, 160000 english vocab and 80000 french vocab = huge model -- could this have been optimized to improve performance, rather than GPUs?

Most importantly, we demonstrated that a simple, straightforward and a relatively unoptimized approach can outperform a mature SMT system, so further work will likely lead to even greater translation accuracies. These results suggest that our approach will likely do well on other challenging sequence to sequence problems.

- Which GPU did you use? (I like to know)

The Nitpicks

- I prefer the related work sections up front - helps to know where we are coming from.
- Page 6, figures split paragraph awkwardly
- Training Details - thorough, but I really wish people would say what didn't work in their papers.
- Lots of redundancies
 - “Deep Neural Networks (DNNs) are powerful models that have”
 - “Deep Neural Networks (DNNs) are extremely powerful machine learning models”
 - “DNNs can only be applied to problems whose inputs and targets can be sensibly encoded with vectors of fixed dimensionality.”
 - “Sequences pose a challenge for DNNs because they require that the dimensionality of the inputs and outputs is known and fixed.”

The Nitpicks

- Typo

3 Experiments

USED

We applied our method to the WMT'14 English to French MT task in two ways. / We used it to directly translate the input sentence without using a reference SMT system and we it to rescore the n-best lists of an SMT baseline. We report the accuracy of these translation methods, present sample translations, and visualize the resulting sentence representation.