***Code*-Switch**

jchuter@stanford.edu

TASK

Neural machine translation of text from informal to formal style.

Structured learning with various modern approaches.

Applicable to tasks where text style should be tailored to the audience.

DATASET

GYAFC Dataset [1]:
The dataset includes ~110,000 pairs of formal and informal sentences. Informal sentences were selected from a subset of Yahoo Answers and paired with formal rewrites from Mechanical Turk.

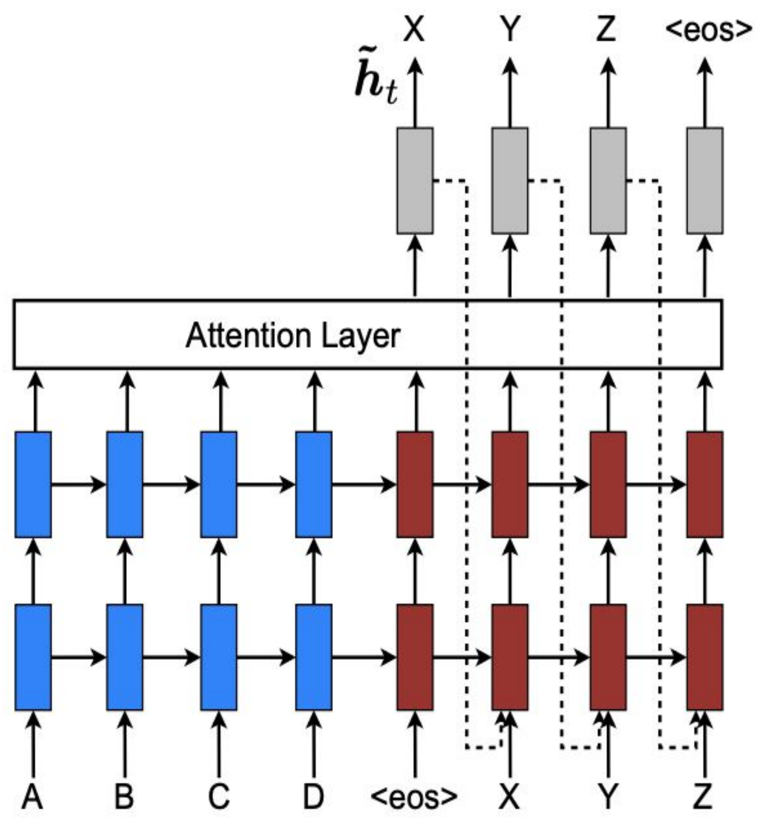
Split into train, eval, and test sets of sizes 104,000, 3,000, 3,000 respectively, where each set samples uniformly from each informal domain (Entertainment & Music, Family & Relationships).

informal	**formal**
Because there is no such thing!!!!!!!!!	It does not exist.
IT DOESN'T MEAN THAT YOU ARE SLUTTY.	It does not mean that you are a loose woman.
i would go out there and kick your butt right now, but i can't swim.	I would travel to that location and physically assault you at this very moment, however, I am unable to swim.

METHODS

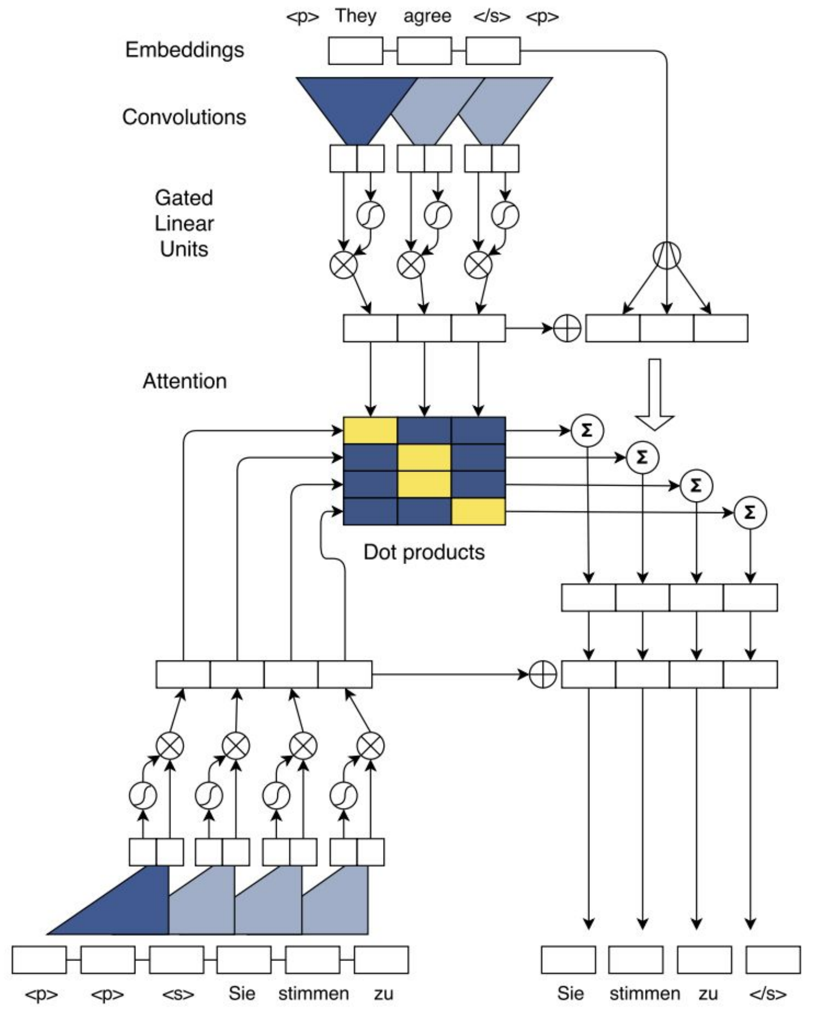
RNN

4 layer neural network, with 2 encoding and 2 decoding layers [2]. Each layer is an LSTM cell with 500 hidden units.



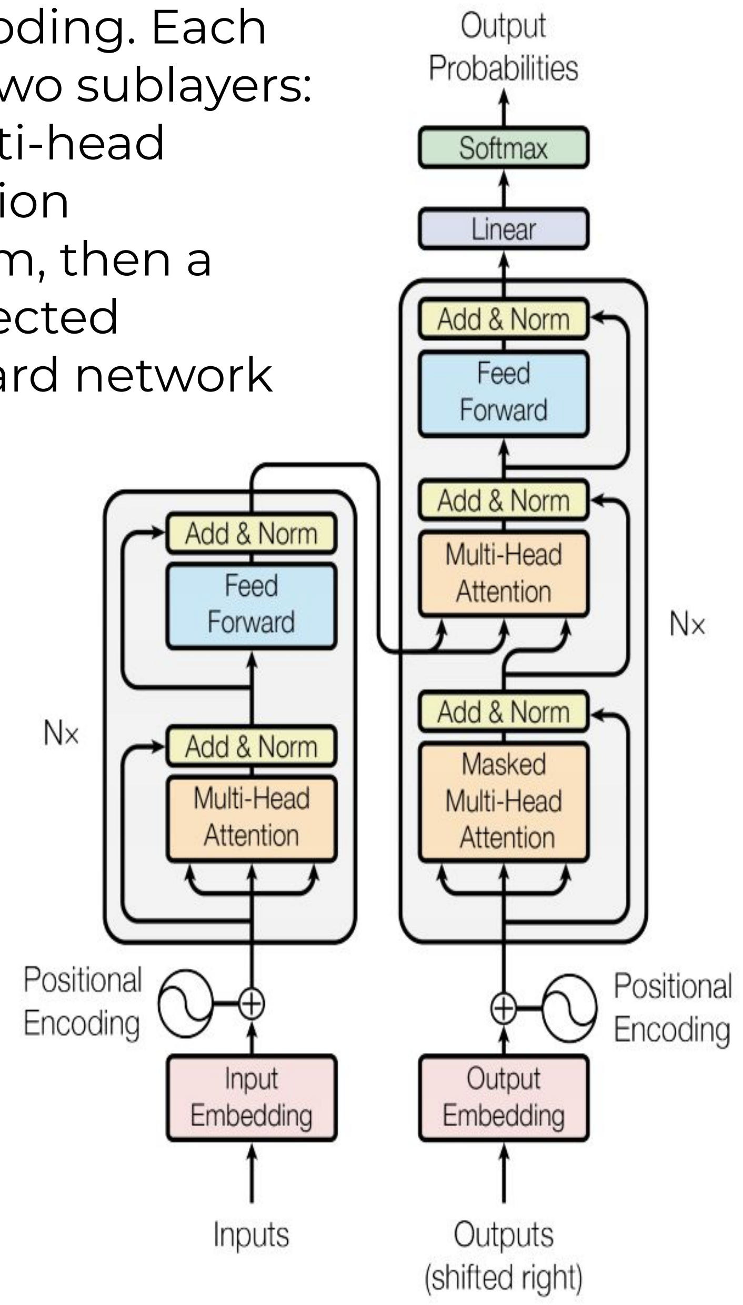
CNN

A CNN which uses gated linear units for easier gradient propagation and attention modules on decoder layers [3]. This specific model uses 2 layers with size 3 kernels.



TRANSFORMER

This transformer has 12 total layers, 6 encoding and 6 decoding. Each layer has two sublayers: first a multi-head self-attention mechanism, then a fully connected feed-forward network [4].



RESULTS

model	Accuracy	Perplexity	Time (s)
RNN	47.60	26.79	3680
CNN	33.1058	86.92	10562
TFMR	44.55	116.32	43200

DISCUSSION

The RNN model trained fastest and achieved best performance.

The results of informal input to formal output have largely passable semantics, fluency, and appropriate style.

input	RNN	CNN	Transformer
Cant touch this.	I cannot touch this.	I cannot touch this.	I cannot touch this.
Hey how's it going?	Hello, how is it going?	Hey, how it goes.	How is it going?
Lets chill.	Let us rephrase to you.	I am not an important thing to do something that way.	Let us keep in mind that we should give us some space.

IMPLEMENTATION

Implemented with OpenNMT [5].

Trained on GCP w/ Tesla-k80 Nvidia GPU; 20,000 steps each model .

FUTURE WORK

Some inputs yield subpar results, due to no similar examples in dataset.

The GYAFC dataset is insufficiently large to train the entire translation task. Multiple solutions could address this.

- Dataset augmentation with similar corpora
- Back-translation [6]
- Transfer learning with pre-trained word embeddings such as GloVE or BERT.

REFERENCES

[1] *Dear Sir or Madam, May I Introduce the GYAFC Dataset: Corpus, Benchmarks and Metrics for Formality Style Transfer.* Tetreault et al. 2018. https://arxiv.org/pdf/1803.06535.pdf

[2] *Effective Approaches to Attention-based Neural Machine Translation*. Luong et al. 2015. https://arxiv.org/pdf/1508.04025.pdf

[3] *Convolutional Sequence to Sequence Learning.* Gehring et al. 2017. https://arxiv.org/pdf/1705.03122.pdf

[4] *Attention Is All You Need* . Vaswani et al. 2017. https://arxiv.org/pdf/1706.03762.pdf

[5] https://github.com/OpenNMT/OpenNMT-py

[6] *Improving Neural Machine Translation Models with Monolingual Data*. Sennrich et al. https://arxiv.org/pdf/1511.06709.pdf