

Email Autoresponder using a Sequence-to-sequence Architecture with BERT

Claudio Scheer
Jose Fernando Possebon Junior



Agenda

01

Motivations & Goals

02

Related Work

03

Proposal

04

Experiment & Results

05

References



Motivations & Goals

- Implement a neural network to solve a specific problem of non-structured data.
- Emails are good source of non-structured data and that are some automation opportunities.
- Spam is just one of issues addressed about emails.
- Filter emails that requires human intervention from what can be automated.



A possible opportunity

A salesperson that sells software licenses receive requests for quotations from their customers about the price of software licenses and payment conditions.

The answers for those emails requests follow some patterns and usually involve checking the customer name, the software required, the number of licenses, and the customer's current price list.



Related Work

- Efficient Natural Language Response Suggestion for Smart Reply
 - Henderson et al. 2017] Henderson, M.; Al-Rfou, R.; Strophe,B.; Sung, Y.-H.; Luk´acs, L.; Guo, R.; Kumar, S.; Miklos, B.;and Kurzweil, R. 2017.
- Smart Reply: Automated Response Suggestion for Email
 - [Kannan et al. 2016] Kannan, A.; Kurach, K.; Ravi, S.; Kauf-mann, T.; Tomkins, A.; Miklos, B.; Corrado, G. S.; Luk´acs,L.; Ganea, M.; Young, P.; and Ramavajjala, V. 2016.
- Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation
 - [Cho et al. 2014] Cho, K.; van Merrienboer, B.; Gulcehre,C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio,Y. 2014.



Proposal

- The first idea was to implement something using the Brazilian Portuguese language.
- There are many more datasets in the English language available, so we decide to go that way.
- We did not implement the piece of checking email inbox and sending the answer email.
- Provide software capable of receiving any text with some context and producing a meaningful answer to that.



Experiment Dataset

- Enron Email Dataset
 - This dataset was collected and prepared by the CALO Project (A Cognitive Assistant that Learns and Organizes). It contains data from about 150 users, mostly senior management of Enron, organized into folders.
 - The corpus contains a total of about 0.5M messages. This data was originally made public, and posted to the web, by the Federal Energy Regulatory Commission during its investigation.
- Parse emails
 - Get only emails with less than 256 characters.
- Final dataset: 40.062 input-target pairs.



Experiment

The Enron Email Dataset

no, i dont mind. : -)

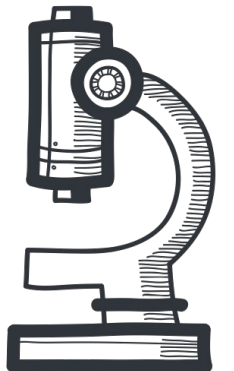
-----Original Message-----

Hi how are you doing? I have a meeting from 4 to 5, do you mind waiting for me?
Thanks.

John

-----Original Message-----

hello



Experiment

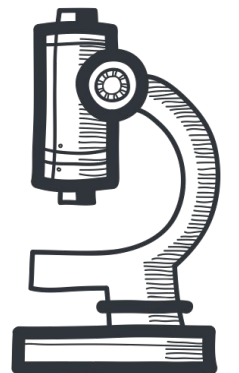
The Enron Email Dataset

yuck yuck

-----Original Message-----

har har

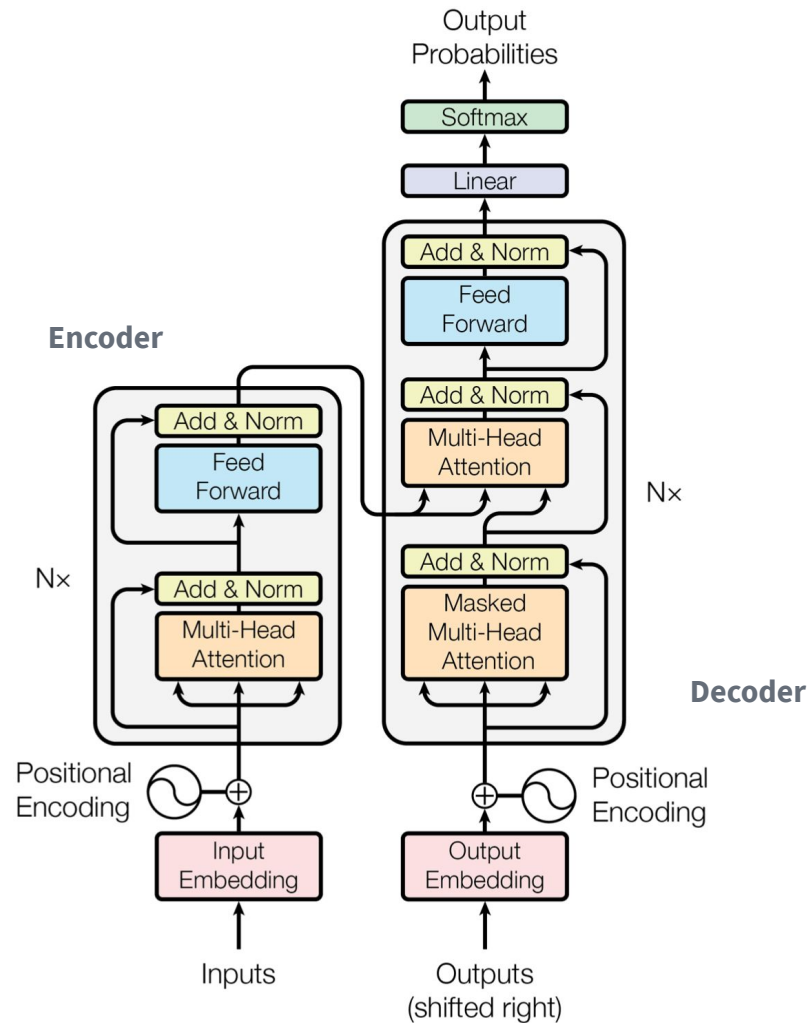
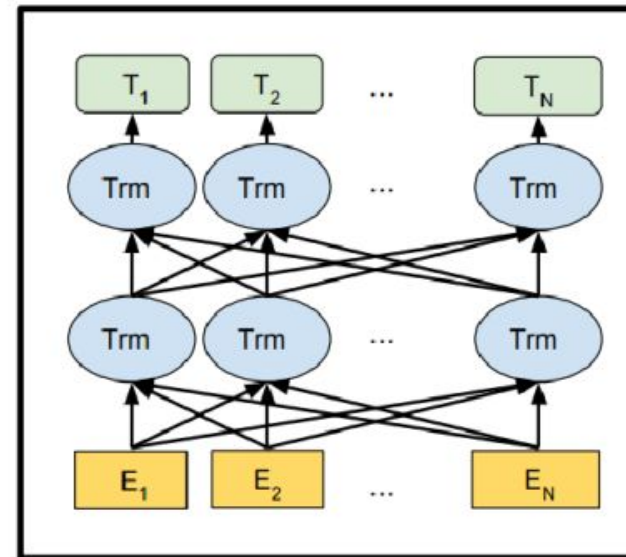
The subject is: **Wine tasting.**



Experiment

BERT

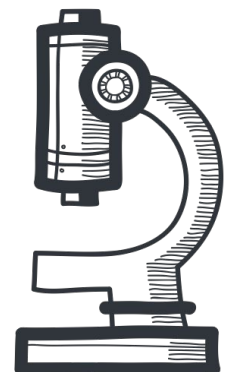
BERT architecture



Transformers architecture

Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	##ing	[SEP]
Token Embeddings	$E_{[CLS]}$	E_{my}	E_{dog}	E_{is}	E_{cute}	$E_{[SEP]}$	E_{he}	E_{likes}	E_{play}	$E_{\#ing}$	$E_{[SEP]}$
Segment Embeddings	E_A	E_A	E_A	E_A	E_A	E_A	E_B	E_B	E_B	E_B	E_B
Position Embeddings	E_0	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}

Words input on BERT



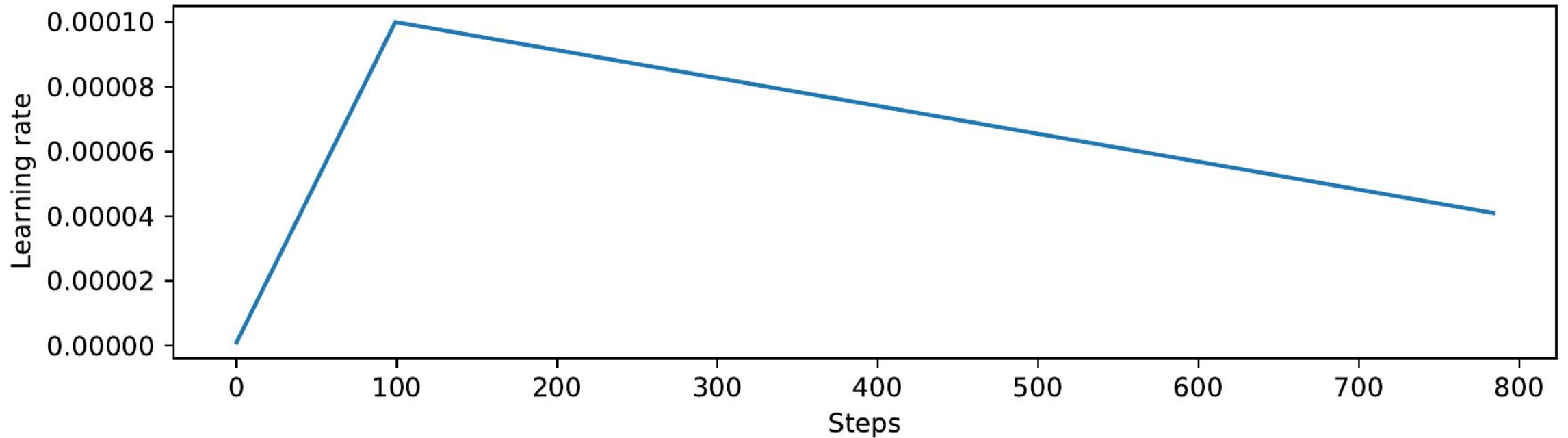
Implementation and Results

Hugging Face

- BERT base;
 - 12 layers;
- Hugging Face and PyTorch;
- NVIDIA T4 GPU on AWS;
 - 16GB;

Learning rate schedule

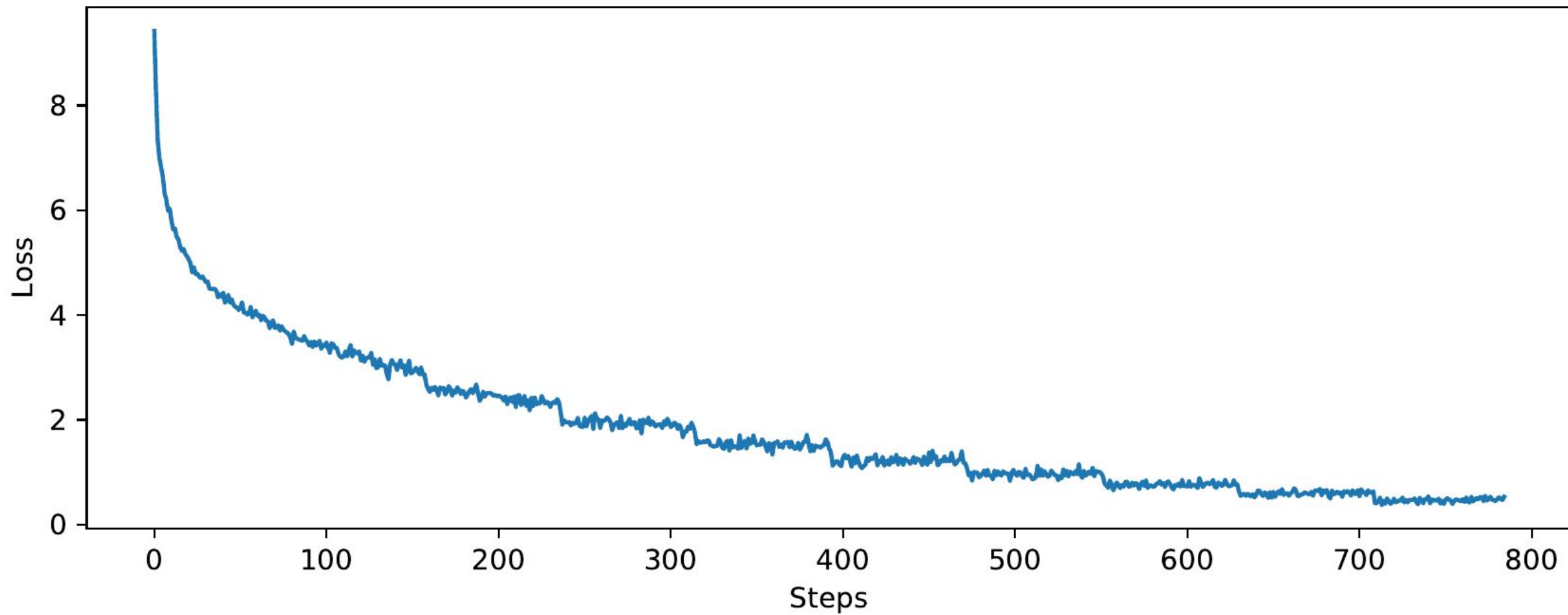
- Warm-up steps: 5000;
- Learning rate: $1e-4$;



Optimizer and hyperparameters

- Adam;
 - Adam epsilon: $1e-4$;
- Batch size: 10;
- Fine-tuning epochs: 10.5;
- Beam search hypothesis: 3;

Loss function



BLEU score

- 21 emails in the evaluation dataset;
- BLEU score = 0.0;

BLEU score

Input	Target	Generated
nevermind. are you at work?	yea	i'm all alone over here...
what re your plans for tomorrow?	drop prentice's car off at shop; go to dome	leaving early tomorrow.
I just refaxed it. For your information, my total hours is 49.	I noted your hours..I will look for the t/s Thanks Lisa!	Thanks Rick, and let's talk after the holidays about a new game plan... Best regards and happy thanksgiving. Jaime Williams.

Generated replies

Email	Reply	AI?	Was it AI?
Let me know if this works.	This works great. Will let you know. as soon as I get a chance, let me know	95.7%	yes
Following are the terms under which I ran the Forward Obligation report and received this message: General SQL error. ORA-00907: missing right parenthesis. Terms: Let me know what you find. Thanks!	Kate, I ran the report with the exact same options and it came right up. Try it again, exactly as before. Usually if it works for one person and not the other it is a security issue. Let me know what happens. Thanks, Brettther	52.2%	yes
Didn't you trade uranium at one time?	Yeah, I know the buisness VERY well.	30.4%	no
Are you free for drinks either Monday or Wednesday?	Yes	69.6%	no
Mons, I would be available on the 25th, 26th or 27th. I cannot make it the week of the 18th. Thanks, Bill.	OK, so, let's see if we can get together later today. I have to leave at 16:00 for a few minutes, but I am sure that I will be out at that moment. Thank you Kim.	39.1%	yes

Other tests

- 1002 emails, 64 epochs
- 10000 emails, 128 epochs (bad)

Miscellaneous

- [Source code](#)
- [Final release \(v0.0.8\)](#)
- [The Enron Email Dataset](#)
- [Dataset parser](#)

References

- [Bahdanau, Cho, and Bengio 2015] Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. **Neural machine translation by jointly learning to align and translate**. In Bengio, Y., and LeCun, Y., eds., 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.
- [Cho et al. 2014] Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. **Learning phrase representations using RNN encoder-decoder for statistical machine translation**. CoRR abs/1406.1078.
- [Devlin et al. 2018] Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2018. **BERT: pre-training of deep bidirectional transformers for language understanding**. CoRR abs/1810.04805.
- [Graves 2013] Graves, A. 2013. **Generating sequences with recurrent neural networks**. CoRR abs/1308.0850.
- [Henderson et al. 2017] Henderson, M.; Al-Rfou, R.; Strope, B.; Sung, Y.-H.; Lukács, L.; Guo, R.; Kumar, S.; Miklos, B.; and Kurzweil, R. 2017. **Efficient natural language response suggestion for smart reply**. ArXiv abs/1705.00652.
- [Kannan et al. 2016] Kannan, A.; Kurach, K.; Ravi, S.; Kaufmann, T.; Tomkins, A.; Miklos, B.; Corrado, G. S.; Lukács, L.; Ganeva, M.; Young, P.; and Ramavajjala, V. 2016. **Smart reply: Automated response suggestion for email**. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [Cho et al. 2014] Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. **Learning phrase representations using rnn encoder-decoder for statistical machine translation**.



Questions



Thank you

Claudio Scheer
Jose Fernando Possebon Junior