

Social Media Chatbot, or, NLP on small data

by Alexander Van Roijen

March 2020

Introduction and Data

Facebook chat data to make a "user"

This presentation will highlight the process of developing a chat bot using a seq2seq model on a small dataset from a facebook messenger chat.

- 51798 messages
- 14000 unique words
- 4 users
- 7 year duration

- Introduction
- Motivation
- Background / Relevant Works
- Methodology
- Results
- Conclusion
- Future Work

Motivation

Why make a chat bot?

- I want more experience in NLP
- Facebook and GroupMe are trying this^[1]
- This is novel in that small data chatbots aren't explored
- I have already done most of the [data parsing and other stuff](#)
- I want to make fun of my friends



Relevant Works and Resources Used

- A Neural Conversational Model
- Attention Boosted Sequential Inference Model
- Many tutorials such as this
- But most importantly this one where I stole all my code from

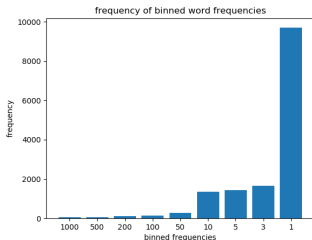
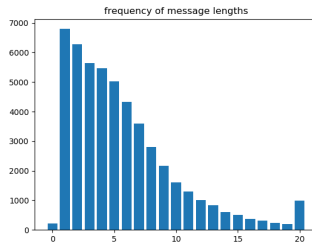
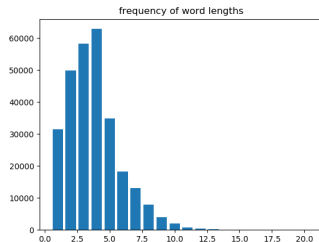
- Data collection
- Processing
- Analysis of distribution
- Create subsets
- Train model
- Analyze
- Reiterate

Where we are so far

- Introduction
- Motivation
- Background / Relevant Works
- Methodology
- Data
- Models
- Results
- Model translation examples
- Conclusion
- Future Work

Data

Data



Thresholds? Typos?

Model Number	dropOut	Epochs	hasUser	seqLen	freqThresh	overFitIndex
1	0.05	500	False	20	10	11
2	0.3	1000	False	20	10	12
3	0.3	1000	True	20	10	12

Table: hasUser = True if data included indication of what user made the statement. freqThresh represents at what threshold we removed a word from the vocabulary. Overfit index is the epoch at which validation loss stopped decreasing.

All used cross entropy loss, ADAM updater, embedding layer of size 128, hidden layer of size 256, consisting of an encoder and decoder utilizing GRUs, and Bahdanau attention matrix [2,3,4].

Results

Model Number	trainLoss	valLoss	optimalValLoss
1	0.26	4.75	2.52
2	1.048	3.53	2.46
3	1.036	3.64	2.49

Table: all loss measured as an average per input per batch of 100 rows.
OptimalValLoss is the validation loss at its lowest point in model training

But is cross entropy loss really what we are after?

Some Good Examples

Model 1

input: be on in a sec

output: ok discord ?

...

input: did you wanna wait for jer at all ? i got some work i wanted to
finish first

output: sure

input: let me know when

output: shoot for 8

Some Bad Examples

Model 1

input: can i get a ride or no ?

output: i gotta play at 800

...

input: how was the first day of work kayla and jerry ?

output: mine was watching a lot of you ont

input: yoooo i saw that keyboard on reddit

output: yeah i was so coming into a way to explain out of a bike to the
final project for

Some Good Examples

Model 2

input: its a hard life

output: yes

input: so three groups ? jendj

output: no volleyball has a ton of groups

input: get rekt

output: space rekt

input: youre free around 5 our time mango ?

output: prob more like 6

input: what do you wanna play ?

output: ova watch

Some Bad Examples

Model 2

input: okay output: i got a scarf existence

...

input: remember i really love this game

output: i take full credit input: alex when you leaving ?

output: were going to be able to leave

input: alright then jer you gonna bike to us

output: are you guys are you guys are you guys

Some Good Examples

Model 3

Input: Jeremy: not too late
gotta be up at 6 : 30

...

input: Kyle: is alex ready ?

output: alex you there ?

input: Jeremy: why would i know who this is

output: i feel like you know the broadcaster

input: Dylan: night

output: night

input: Alex: as in she can come

output: lets go

Some Bad Examples

Model 3

input: Alex: mangos wanna do somtin today ?

output: kyle is kyle

...

input: Kyle: jerry when do you start work ? soon

output: 5 is debating to

input: Kyle: oh shit yeah we were talking about high

output: and i think you still wanna play

input: Dylan: should i restart ?

output: my bright funds

Conclusion

- Data manipulation
 - large impact on performance
 - not an exact science (yet?)
- Cross Entropy Loss functions
 - Optimize correctness of words, not meaning
 - Are the end all of ML
 - Good for machine translation, but no necessarily communication
- Overall
 - Model 2 performs the best in my eyes
 - POS tagging doesnt solve many problems

- Establish better baseline with Naive Bayes and N-Grams
- Use pretrained word embeddings
- Play around with different thresholds and dimensions
- Use cosine similarity for new loss function
- Incorporate other group chats
- Use in depth one on one versus one with many
- Visualize attention matrixes to determine relationships

Questions?

Feel free to checkout my github where I will be posting this model and report. Also email me anytime at ajvroijen@gmail.com

References

- ① <https://help.groupme.com/hc/en-us/articles/115004382087-What-is-Zo->
- ② <https://arxiv.org/pdf/1506.05869.pdf>
- ③ https://www.tensorflow.org/tutorials/text/nmt_with_attention
- ④ <https://arxiv.org/pdf/1409.0473.pdf>
- ⑤ https://en.wikipedia.org/wiki/Instant_messaging
- ⑥ http://nlpprogress.com/english/natural_language_inference.html
- ⑦ <https://www.aclweb.org/anthology/D19-1271.pdf>
- ⑧ <http://ai.dinfo.unifi.it/paolo//ps/tnn-94-gradient.pdflearningtolatchwithdynamicalsystems>