Sangyoon Kim

CPSC393: Machine Learning

Dr. Chelsea Parlett

April 28, 2023

Introduction

One of the main applications of natural language processing (NLP) is text generation, which involves generating natural language text similar to human-written text. This text-generation function has been practiced in a wide range of fields. This experiment describes the process of building a word-level neural language model and using it to generate text. A neural language model is a machine learning model in which a sequence of previous words is trained to predict the probability of a given word. This model works by predicting the next word in order based on the preceding word. This experiment presents a word generation model using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) neural networks based on Moby-Dick text files. This model aims to generate new text similar to the style of the novel Moby-Dick. Accuracy metrics are used to evaluate the performance of the model.

Analysis

The dataset used for this project was a collection of text documents, which were preprocessed to remove punctuation and convert all text to lowercase. The resulting corpus was then split into training, validation, and test sets. The training set was used to train the neural language model, while the validation set was used to tune the hyperparameters of the model. The test set was used to evaluate the performance of the model.

Methods

The dataset used for this project, mobydick.txt, was downloaded from Project Gutenberg. The neural language models used in this project were feed-forward neural networks with embedding layers, then one or more hidden layers, and finally, the SoftMax output layers. The size of the embedding vector space, a parameter specifying the number of dimensions to represent each word, was defined as 50. Use two long short-term memory (LSTM) hidden layers with 100 memory cells. Using more memory cells and deeper networks will result in better results. A high-density, fully connected layer with 100 neurons is connected to the LSTM hidden layer to interpret the features extracted from the sequence. The output layer predicts the next word with a single vector of vocabulary size, with probabilities for each word in the vocabulary. The SoftMax enable

function ensures that the output has the characteristics of a normalized probability. Compiles the model by specifying the categorical cross-entropy loss. An efficient Adam implementation for the mini-batch gradient descent method is used, and the model's accuracy is evaluated. One hundred epochs were used with a batch size of 128 to speed up the training work. A performance summary is displayed at the end of each batch update, including the loss and accuracy evaluated in the training data. In the case of the second model, GRU was used differently from the LSTM of the first model.

Seed sentences were selected to generate text using the trained model, and the following word in the sequence was predicted using the model. Prediction words were added to the end of the sequence, and the process was repeated until the text of the desired length was generated. Each model outputs ten predicting results.

Results

The first model has 2.0041 as loss, and 0.5407 as accuracy.

The outputs from first trial:

time the faint air had become a complete calm so that whether or no the pequod was now fairly entrapped in the smell with no hope of escaping except by its breezing up again issuing from the cabin stubb now called his crew and pulled off for the stranger drawing across her bow he perceived that in accordance with the fanciful french taste the upper part of her stempiece was carved in the likeness of a huge drooping stalk was painted green and for thorns had copper spikes projecting from it here and there the whole terminating in a symmetrical folded

bulb of a bright red colour fast lashed to the smoking and mainmast the ship the signs yellow cuba rolled hence the fourth point is far as darkness came upon authoritative blinds account of his tormented jaw meanwhile ahab was the whale the whale is thrown down upon the side

reclining and shaking his cap the caps turn to jig it now shake their tassels soon now would all the waves were women then go drown and chassee with them evermore naught so sweet on may not match those swift glances of warm wild bosoms in the dance when the overarboring arms hide such ripe bursting grapes sicilian sailor reclining tell me not of it hark ye interlacings of the lip heart hip all graze unceasing touch and go not taste observe ye else come satiety eh pagan nudging tahitan sailor reclining on a mat hail holy nakedness of our dancing

heevaheeva ah low veiled high palmed tahiti i observe that the unconquerable one whale rolls high

up lashed in the rack whence though roused his lampfeeder at the undraped spectacle of the two crippled settlement slicings the parted projecting of the upper part loath to stern stubb was counsel to

sorry remainder nevertheless leviathan is of so mighty a magnitude all his proportions are so stately that the same deficiency which in the sculptured jove were hideous in him is no blemish at all nay it is an added grandeur a nose to the whale would have been impertinent as on your physiognomical voyage you sail round his vast head in your jollyboat your noble conceptions of him are never insulted by the reflection that he has a nose to be pulled a pestilent conceit which so often will insist upon obtruding even when beholding the mightiest royal beadle on his

throne the lighted deliberately lowered the buttend the six tower in length leo the wildness of the sciences and all sprang upon the midnight waves entitled standing in his erect attitude ahab and the replenished compass rolled in the sea and the the watery whale of a solecism and learned

and then independent hilarious little flask enters king presence in the character of abjectus or the slave it is not the least among the strange things bred by the intense artificialness of seausages that while in the open air of the deck some officers will upon provocation bear themselves boldly and defyingly enough towards their commander yet ten to one let those very officers the next moment go down to their customary dinner in that same cabin and straightway their inoffensive not to say deprecatory and humble air towards him as he sits at the head of the table this is

marvellous gentlemen i know something here and i may float but neither rakes of julys and profit here that on the smallest coast it is shown but the elephant is not a rare thing when that could not be able to have a sort of honor in the same time

the thought that in pious bildad might be found some salvation spite of his seven hundred and seventyseventh lay when i felt a sudden sharp poke in my rear and turning round was horrified at the apparition of captain peleg in the act of withdrawing his leg from my immediate vicinity that was my first kick that the way they heave in the marchant he roared thou sheephead spring and break thy backbone why ye spring i say all of quohog spring thou chap with the red whiskers spring there scotchcap spring thou green pants spring i say all of ye

have a dangerous economical feeling on japan and harpooneers and know thee how that this coin you relent that i must say but i have heard of dat of his breeding nevertheless this soothing old barrels he pushed open down and exhibited aye will ye be invisibly and uninterpenetratingly wants

so crossed and recrossed and in a thousand ways entangled the slack of the three lines now fast to him that they foreshortened and of themselves warped the devoted boats towards the planted irons

in him though now for a moment the whale drew aside a little as if to rally for a more tremendous charge seizing that opportunity ahab first paid out more line and then was rapidly hauling and jerking in upon it that way to disencumber it of some sight more savage than the embattled teeth of sharks caught and in the mazes of the line loose harpoons

and lances with their undoubted superiority over the constant vision of man which is countersunk into the summit of the whale and vehement puff after puff from the cap of the topgallantmast so that the line passed forward on deck cried stubb matters afterwards started in the rigging is not

here and there a candle like a candle moving about in a tomb at this hour of the night of the last day of the week that quarter of the town proved all but deserted but presently i came to a smoky light proceeding from a low wide building the door of which stood invitingly open it had a careless look as if it were meant for the uses of the public so entering the first thing i did was to stumble over an ashbox in the porch ha thought i ha as the flying particles almost choked me are these

ashes on board and something on the deck and with a jackknife gently disasters and every limbs where noise months hit to show the tremendous burden of a valorruined man nor would forgot or fear of it quicker than a rather life where all the past voice between these mighty

visitors with a bunch of keys at his side sir clifford thinks of charging twopence for a peep at the whispering gallery in the spinal column threepence to hear the echo in the hollow of his cerebellum and sixpence for the unrivalled view from his forehead the skeleton dimensions i shall now proceed to set down are copied verbatim from my right arm where i had them tattooed as in my wild wanderings at that period there was no other secure way of preserving such valuable statistics but as i was crowded for space and wished the other parts of my

body you are really sailing in him and then grasping the most satisfactory he nor has no selfesteem and no caps at all events i wear a worthy lima lies up to the door and murmured ho anything heaven ha truth is a deacon marvellous not i never asked the

sixteen entire hours must elapse before i could hope for a resurrection sixteen hours in bed the small of my back ached to think of it and it was so light too the sun shining in at the window and a great rattling of coaches in the streets and the sound of gay voices all over the house i felt worse and last i got up dressed and softly going down in my stockinged feet sought out my stepmother and suddenly threw myself at her feet beseeching her as a particular favour to give me a good slippering for my misbehaviour

anything indeed this is the life of my own proper harpoon with a white whale of a greenland whale

than a little brains of view did not think of any gorings of try somewhat humble them out on thee as the guernseyman looked about the isle of desolation go and

some fathoms of it taken to the windlass and stretched to a great tension pressing his foot upon it till the rope hummed like a harpstring then eagerly bending over it and seeing no strandings ahab exclaimed and now for the at one extremity the rope was unstranded and the separate spread yarns were all braided and woven round the socket of the harpoon the pole was then driven hard up into the socket from the lower end the rope was traced halfway along the length and firmly secured so with intertwistings of twine this done pole iron and the three

inseparable and the right whale the bleak rustlings of the sea had been made that infinite from the midday sea in the estimation of a clumsy temple the keels seemed majestic of things both his even were stripped the english element and pods of the fancied silence and carried it

The second model has 2.1512 as loss, and 0.5352 as accuracy.

The output of the second model:

we have elsewhere his dead back in a rolling sea and finally descend into the gloom of the hold and bitterly sweating all day in that subterraneous confinement resolutely manhandle the clumsiest casks and see to their stowage to be short among whalemen the harpooneers are the holders so called poor queequeg when the ship was about half disembowelled you should have stooped over the hatchway and peered down upon him there where stripped to his woollen drawers the tattooed savage was crawling about amid that dampness and slime like a green spotted lizard at the bottom of a well and

the walrus is answered it he is an impartial volume but not on the sea ringbolts and such a cosy valley of respect for me and all struggling towards there gomorrah felt him with the land there was far to the mainmast to the mutual a the breathless pod till

of ye nor can the oldest nantucketer thus ends book i folio and now begins book ii octavo embrace the whales of middling magnitude among which present may be the grampus ii the black fish iii the narwhale iv the thrasher v the killer why this book of whales is not denominated the quarto is very plain because while the whales of this order though smaller than those of the former order nevertheless retain a proportionate likeness to them in figure yet the quarto volume in its dimensioned form does not preserve the shape of the folio volume but the octavo

isles of the samuel coat of this of pillars his brow of thick air found as in mankind ob no more cool there is game to have done natural at the whale the captain but scarcely entering it was an folios by he was gradually more to temper that blooddripping

his brain slowly crossing the deck from the scuttle ahab leaned over the side and watched how his shadow in the water sank and sank to his gaze the more and the more that he strove to pierce the profundity but the lovely aromas in that enchanted air did at last seem to dispel for a moment the cankerous thing in his soul that glad happy air that winsome sky did at last stroke and caress him the stepmother world so long threw affectionate arms round his stubborn neck and did seem to joyously sob over him as if over one

of his uplifted eyes and wetter curled very a other as you will not touch ineffable expand for in his bottom dust in it the wind could be shortened but some twenty fathoms the horrors which shows their general presumption have perceive these manifold or beasts of stacked each years

england rocks on the seacoast which agassiz imagines to bear the marks of violent scraping contact with vast floating should say that those rocks must not a little resemble the sperm whale in this particular it also seems to me that such scratches in the whale are probably made by hostile contact with other whales for i have most remarked them in the large fullgrown bulls of the species a word or two more concerning this matter of the skin or blubber of the whale it has already been said that it is stript from him in long pieces called blanketpieces

like exceeding mout the man is very service that shipkeepers were quite a mighty start bread during the ship all heartwoes a with that in vain load and temporary builders threepence in a panther times of wood and two ships since is not the pagan scraps and might be found

sudden humor assisted memory by snatching him up bodily and thrusting his head into a great empty wooden trencher while tashtego knife in hand began laying out the circle preliminary to scalping him he was naturally a very nervous shuddering sort of little fellow this breadfaced steward the progeny of a bankrupt baker and a hospital nurse and what with the standing spectacle of the black terrific ahab and the periodical tumultuous visitations of these three savages whole life was one continual lipquiver commonly after seeing the harpooneers furnished with all things they demanded he would escape from their clutches into

his little pantry adjoining and fearfully thigh to your hammock all the deck for the see till all comparison and the drugg was narrated the names with hard eyes dines upon the cabin but also of his cap out to it my first i i do but would see have

mouth to be oppositely voided by the gaping wound nor was this all it was unsafe to meddle with

the corpses and ghosts of these creatures a sort of generic or pantheistic vitality seemed to lurk in their very joints and bones after what might be called the individual life had departed killed and hoisted on deck for the sake of his skin one of these sharks almost took poor hand off when he tried to shut down the dead lid of his murderous jaw the whalingspade used for cuttingin is made of the very best steel is about the bigness

of a spread hand and so sure at last the most coalblack statement tistig as it is a head of his first involutions of their book of the haunts accidents even the whale with the garden of the afternoon of the plagues of saying fail but symbolically cut thus cannot

arms with measureless imprecations he shouted out aye and chase him round good hope and round the horn and round the norway maelstrom and round flames before i give him up and this is what ye have shipped for men to chase that white whale on both sides of land and over all sides of earth till he spouts black blood and rolls fin out what say ye men will ye splice hands on it now i think ye do look shouted the harpooneers and seamen running closer to the excited old man sharp eye for the white whale a sharp

lance for god vibrated almost to the while considerations and pillaged with his business the whale was not but a right whale passed the fiend and all the ship and all asia from his voice the two indifference lifting it and in the chapel we recognised a bloodvessels in the

shoot at them from the deck owing to the resistance of the water but to shoot down upon them is a very different thing now it was plainly a labor of love for captain sleet to describe as he does all the little detailed conveniences of his but though he so enlarges upon many of these and though he treats us to a very scientific account of his experiments in this with a small compass he kept there for the purpose of counteracting the errors resulting from what is called the of all binnacle magnets an error ascribable to the horizontal

pen loath his plain and then if me mine dost thou be a boats between day who had been ninety drawers they were charing and leak says was at the chased drawn up merrymad off ashore be humorously a bedford had i i perceived that stubb placed along with the

the fishermen will seldom give chase to one of these grand turks for these grand turks are too lavish of their strength and hence their unctuousness is small as for the sons and the daughters they beget why those sons and daughters must take care of themselves at least with only the maternal help for like certain other omnivorous roving lovers that might be named my lord whale has no taste for the nursery however much for the bower and so being a great traveller he leaves his anonymous babies all over the world every baby an exotic in good time

and surround each not this moment ahab was not a swaths of grapes darting suddenly i offer plant to a burst overhead and whose event is been derived into their thousands of a sea there will not fail speechless the first australian barrens the wind was seen the names of

be sober scientific delineations by those who know in old collection of voyages there are some plates of whales extracted from a dutch book of voyages ad entitled whaling voyage to spitzbergen in the ship jonas in the whale peter peterson of friesland in one of those plates the whales like great rafts of logs are represented lying among iceisles with white bears running over their living backs in another plate the prodigious blunder is made of representing the whale with perpendicular flukes then again there is an imposing quarto written by one captain colnett a post captain in the english

navy entitled voyage berry then swear i would fain let me take that considering them that is answer is a sudden terrors but filled among their clenched and all a physeter or response mate seldom as toil were children crossed in the tertiary muskiness he cannot be seen if i

Discussion

Overall, the accuracy of predicting the next word in the sequence is just over 50%. The loss of the first tested model was 2.00 and the accrual was 54%. The second model tested had a loss of 2.15 and an accrual of 53%. The first model has a lower loss but a higher accrual. By a narrow margin, the first model yields better results. It suggests that the LSTM model was more fit than the GRU model in this case. However, the choice between LSTM and GRU depends on the specific task and dataset, and both models have their own strengths and weaknesses in terms of functionality. The generated text was evaluated qualitatively, and it was found to be coherent and grammatically correct, although it sometimes lacked semantic coherence.

Both LSTM and GRU are types of recurrent neural networks designed to address the vanishing gradient problem in RNN. LSTM has a more complex architecture with three gates that control the flow of information in and out of the cell state. This allows the model to selectively forget or remember information from previous time steps, making it effective for a wide range of natural language processing tasks. However, it also requires more computations and is generally slower and more computationally expensive. On the other hand, GRU has a simpler architecture with two gates that combine the memory and hidden state, making it faster and more memory efficient than LSTM. While it may be less effective than LSTM for some tasks, recent studies have shown that it can achieve comparable or even better performance in certain cases, especially when dealing with shorter sequences.

Reflection

In the process of building and training the word-level neural language model, several challenges were encountered, including selecting appropriate hyperparameters, choosing a suitable pre-trained word embedding model, and dealing with issues such as overfitting and vanishing gradients. However, by carefully tuning the model and using best practices for training neural networks, it was possible to build a model that was able to generate coherent and grammatically correct text. In the future, it would be interesting to explore more advanced techniques such as recurrent neural networks and attention-based models, which are known to be particularly effective for natural language processing tasks.