CSCM38: Adv Topic - Artificial Intelligence and Cyber Security - Coursework 2: Comparing an RNNs LSTM and GRU Cells

Andy Gray
445348

18/12/2020

1 Introduction

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

We will be exploring the proposed solution and the libraries, dataset, preprocessing, algorithms, metrics that will get used for comparison and the NN parameters that we used. We will then be analysing the results, and discussing what insights they provide, ending with a conclusion.

2 Proposed Method

We will be creating an experiment that will compare the two different types of RNN cells, the LSTM and the GRU. We used several parameters to train and test the different RNNs as well as several other metrics to be able to compare the performance of the cells. The RNNs aim to be able to accurately predict if a Tweet posted on Twitter is about a real disaster or not. This experiment used many different Python 3 libraries.

2.1 Libraries & Frameworks

We used a collection of different Python 3 libraries to conduct this experiment. We used Tensorflow 2 [1] for creating the RNN model, LSTM and GRU cells. It got also used for preprocessing the text and sequence with 'Tokenizer' and 'Pad Sequence'. We used Sci-Kit Learn [3] for splitting the dataset and for creating the confusion matrix. Additionally, Pandas [] got used for handling the dataset and Numpy to allow the other libraries to be able to do their scientific calculations. We also used NLTK [] for NLP's stop words as well as some of Python's extra libraries os, time, re and string.

2.2 Dataset and Preprocessing

We used a dataset from Kaggle called "Natural Language Processing with Disaster Tweets" [2]. While a training and CSV file is available, we decided to use the training set due to the test dataset not having any labels. Therefore if we used this dataset, it would be hard to compare how well each cell performed on with an unseen dataset.

The dataset had a shape of 7613, 5. These include the features 'id', 'keyword', 'location', 'text' and 'target'. Due to the 'id' feature not having any relevance and the 'keyword' and 'location' containing null values, these features got dropped from the dataset. The dataset's targets were either a 0 for a non-disaster tweet or 1 for a disaster. There were 4342 non-disaster and 3271 disaster observations. We then removed the characters **[List the characters]** from the dataset's text. We have done this to make sure that the RNN focuses on the contents of the text and not have to focus on the punctuation as this could impact on the model's performance if we had left them in. We also removed all the stop words from the text by using the NLTK library. This action got done to make sure that these stop words also don't impact on the model's understanding of the text. The stop words included: [list stop words]. Along with removing the stop words, we removed any URLs that were in the tweets as these have no relevance on if the tweet is about a disaster or not. When the stop words, punctuation and URLs got removed from the text, we ended up with 17,971 different words contained within all the tweets. The most common words appearing were 'like' (345), 'I'm' (299), 'amp' (298), 'fire' (250) and finally 'get' (229).

Once the dataset was all preprocessed, we split the data set into 2/3 training 1/3 testing. We also then split the training set into an 80/20 split of training and validation data. We did this to see if the dataset was getting overfitted within its training. Additionally, we did this to add an extra method of comparison between the cells. We then tokenised the unique words to create a word index to give the text a number representation for feeding through the RNNs. We then padded the sentence to 20 sequences. We did this to ensure that the length of text within the text would all match up, as the tweets have varying sizes anyway and with the previous clean up actions being down, potentially additional text has also been removed.

An example disaster text is "malaysia airlines flight 370 disappeared 17months ago debris found south indian ocean" and a non-disaster is "walk plank sinking ship".

2.3 Algorithms

For this experiment, we will use the LSTM and the GRU cells. These are both modifications of the RNN. The vanilla RNN, an unaltered RNN, is a robust network. However, it suffers from some issues. These issues are that it only has a short term memory, a suffers from a vanishing gradient point and an exploding gradient too.

2.3.1 LSTM

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

2.3.2 GRU

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

2.4 Metrics for Comparison

To be able to compare the model's performances, we used several metrics. These metrics include the time it took to train the model, the model's accuracy and loss, along with a validation loss and accuracy. We used the time to see how long it would take to train the model. We only timed how long the models took to train from the moment we ran the fit method and not with the initial set up the parameters. This action was to allow us to be able to see how long exactly the model took to complete the number of epochs set and finishing training. We used the accuracy, loss and validation metrics as additional information to see if the model was overfitting.

We also used a confusion matrix to be able to see how well the different models performed on an unseen dataset. Using a confusion matrix allowed us to see if the model was biased in a particular direction or if it was predicting well. Additionally, this allowed another form of comparison to see how accurate the models were on the unseen testing data.

2.5 RNN Parameters

For both the LSTM and GRU RNNs, we used a set value for the parameters. We made sure that these stayed the same when training both models to see which one performed better given each situation.

The loss function used was the binary cross-entropy, and the optimiser was adam with a learning rate of 0.001. These we both implemented using TensorFlows methods. We decided to keep everything within TensorFlow to make sure that everything was consistent and that having these implemented by SKLearn could have been implemented differently, resulting in changes to the model's outputs that are not as a result of the models themself. Therefore, as a result, creating a potential variation that would take focus away from what we were trying to compare.

The metric selected in the model's initiation was the accuracy metric, and we used 20 epochs for training. However, we did use a total of three different cells for each model. These were 32, 64 and 128 cells. We have done this to see how these affect the models training speed and performance in predicting.

Each LSTM and GRU also had an embedding layer of the number of unique words, by 32 and input length of the maximum size. Additionally, all variations had a dense output layer with an activation function of sigmoid and 1 neuron. The variations to the RNNs came in the hidden layers. One version had only 2 hidden layers while another had 4 layers. Each hidden layer had a dropout rate of 0.1. For the RNNs with 2 hidden layers, the return sequence was false, and for the 4 hidden layers, this was true.

3 Results and Discussion

When we compare the direct metrics that were produced by the model's while they were training (see table 1), we can see that the LSTM with 32 cells trained the fastest with an average time of 30.2 seconds while the most prolonged time training was the deep LSTM with 147.456 seconds. A fascinating insight is that the GRU became quicker when we added more layers and cells and then beat the LSTM when we had 128 cells. The GRU, with 128 cells beat the LSTM by 2.476 seconds, and the deep GRU beat the Deep LSTM by 39.294 seconds.

Number of Cells	LSTM	GRU	Deep LSTM	Deep GRU
32	30.2 secs	33.608 secs	81.772 secs	84.288 secs
64	35.876 secs	36.504 secs	96.626 secs	103.72 secs
128	44.892 secs	42.416 secs	147.456 secs	108.162 secs

Table 1: The RNNs average training times.

When we look at the model's metric results for 32 cells (see table 2, we can see that deep GRU had the best scores for all of the metrics apart from the loss and validation loss. However, when we compare the prediction accuracy to the LSTM's scores, we can see that the LSTM had only 0.006% less but took 63.27 seconds less to train and all its other metrics were very close to the deep GRU. Therefore, showing that there is not much difference in the predicted output compared to the overall training time.

Metric	LSTM	GRU	Deep LSTM	Deep GRU
Loss	0.021	0.020	0.026	0.021
Accuracy	0.987	0.987	0.987	0.988
Validation Loss	1.564	1.279	1.406	1.494
Validation Accuracy	0.761	0.763	0.763	0.767
Prediction Accuracy	0.770	0.767	0.763	0.776

Table 2: 32 cells training metric results.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged.

Metric	LSTM	GRU	Deep LSTM	Deep GRU
Loss	0.024	0.020	0.024	0.022
Accuracy	0.987	0.988	0.987	0.988
Validation Loss	1.558	1.584	1.579	1.541
Validation Accuracy	0.755	0.766	0.765	0.762
Prediction Accuracy	0.766	0.766	0.770	0.766

Table 3: 64 cells training metric results.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged.

Metric	LSTM	GRU	Deep LSTM	Deep GRU
Loss	0.028	0.022	0.029	0.021
Accuracy	0.986	0.987	0.986	0.987
Validation Loss	1.450	1.445	1.398	1.539
Validation Accuracy	0.759	0.764	0.770	0.758
Prediction Accuracy	0.766	0.771	0.767	0.763

Table 4: 128 cells training metric results.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.

LSTM	119	330	**GRU**	**1175**	314
	247	737		**271**	753
Deep LSTM	1191	339	**Deep GRU**	**1236**	327
	255	728		**210**	714

Table 5: 32 cells confusion matrix results.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since

LSTM	1167	310	**GRU**	**1174**	315
	279	757		**272**	752
Deep LSTM	1208	340	**Deep GRU**	**1182**	324
	238	727		**264**	743

Table 6: 64 cells confusion matrix results.

LSTM	1178 TP	320 FN	**GRU**	**1173**	304
	268 FP	747 TN		**273**	763
Deep LSTM	1191	331	**Deep GRU**	**1163**	313
	255	736		**283**	754

Table 7: 128 cells confusion matrix results.

the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

References

[1] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] Kaggle. Real or not? nlp with disaster tweets, 2020.

[3] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.