

Investigation of the Batch Size Influence on the Quality of Text Generation by the SeqGAN Neural Network

Nikolay Krivosheev¹, Ksenia Vik^{1,2}, Yulia Ivanova¹ and Vladimir Spitsyn¹

¹ Tomsk Polytechnic University, Lenin Avenue, 30, Tomsk, 634050, Russia

² Tomsk State University of Architecture and Building, Solyanaya square, 2, Tomsk, 634003, Russia

Abstract

One of the problems of text generation using the LSTM neural network is a decrease in the quality of generation with an increase in the length of the generated text. There are various solutions to improve the quality of text generation based on generative adversarial neural networks. This work uses preliminary training of the LSTM neural network based on the MLE approach and further training based on the SeqGAN neural network. Based on the presented results, we can conclude that the SeqGAN-based approach allows to increase the quality of text generation according to the NLL and BLEU metrics. The study of the influence of the batch size, in the process of competitive training of the SeqGAN neural network, on the quality of text generation has been carried out. It is shown that with an increase in the batch size, in the process of adversarial learning, the quality of LSTM neural network training increases. In this work, the Monte Carlo algorithm is not used in the training process of the SeqGAN neural network. For training and testing algorithms, image captions from the COCO Image Captions data sample are used. The quality of text generation based on the NLL and BLEU metrics has been assessed. Examples of the results of generating texts with an assessment of the quality of examples according to the BLEU metric are given.

Keywords

SeqGAN, text generation, adversarial learning, reinforcement learning, batch size

1. Introduction

Algorithms for automatic text generation are in demand and are widely used in various tasks, such as: automatic generation of random texts, automatic translation of texts, text summarization, generation of captions for images.

One of the problems of text generation using the LSTM [1] neural network is a decrease in the quality of generation with an increase in the length of the generated text. There are various solutions to improve the quality of text generation based on generative adversarial neural networks. In this work, we use preliminary training of the LSTM neural network based on the MLE [2] approach and further training based on the SeqGAN [3] neural network. The training process based on the SeqGAN neural network does not use the Monte Carlo algorithm proposed in paper [3]. The study of the influence of the batch size on the quality of LSTM neural network training in the process of adversarial training based on the SeqGAN neural network was carried out. It is shown that with an increase in the batch size, in the process of adversarial learning, the quality of LSTM neural network training increases. Also, an increase in the batch size leads to an increase in the training time of the algorithm.

For training and testing algorithms, captions to images from the COCO Image Captions [4] data sample are used. The proposed sample contains captions in English. In this work, word-by-word text generation is used. The maximum sentence length is 20 words; most sentences in this set contain about

GraphiCon 2021: 31st International Conference on Computer Graphics and Vision, September 27-30, 2021, Nizhny Novgorod, Russia

EMAIL: nikola0212@mail.ru (N. Krivosheev); kvv11@tpu.ru (K. Vik); jbolotova@tpu.ru (Y. Ivanova); spvg@tpu.ru (V. Spitsyn)

ORCID: 0000-0002-1874-6973 (N. Krivosheev); 0000-0001-8706-4290 (K. Vik); 0000-0002-8161-3575 (Y. Ivanova); 0000-0001-5978-1321

(V. Spitsyn)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

10 words. COCO Image Captions sample data is taken from analogue [5]. The data is posted on the website [6]. Thus, the preprocessing of the data selection coincides with that implemented in [5].

The testing process uses a pretrained neural network based on the MLE approach. In the MLE-based pre-learning process, the batch size is 1,000 examples and does not change. Next, we tested the LSTM neural network trained on the basis of the SeqGAN approach using various batch sizes. Tested on a batch size of 40, 400 and 4000 samples.

In this work, the software implementation of the investigated approaches in the Python language was carried out. The software implementation is presented on the site [7] and is based on the work [8].

2. Description of the used data sample

In this work, we used the COCO Image Captions (Common Objects in Context) dataset. This set consists of 330000 images, of which 220000 samples are marked. All images are accompanied by annotations stored in json format. This set has various types of annotation, such as:

- object detection;
- keypoint detection;
- stuff segmentation;
- panoptic segmentation;
- denepose;
- image captioning.

To form a textual data sample, captions to images in English from the COCO Image Captions sample are used. The maximum length for an example is 20 words. Most of the sample texts contain about 10 words. The data sample was taken from the source [5], so its preprocessing corresponds to that presented in article [7]. Words mentioned less than 10 times have been removed from the sample, examples containing these words have also been removed. The number of words used in the dictionary is 4837 words. This dataset contains 80000 examples in the training set and 5000 examples in the test set.

3. Description of the influence of batch size on the SeqGAN learning process

In this work, we use the SeqGAN neural network proposed in paper [3]. Our work does not use the Monte Carlo algorithm proposed in work [3]. One of the hyperparameters affecting the training process of the SeqGAN neural network is the batch size. This parameter affects the number of examples generated by the neural network by the generator and further evaluated by the neural network by the discriminator. An example of text evaluation by a neural network by a discriminator is shown in the image Figure 1.

Examples of texts in the package						Text quality
A	young	girl	sitting	on	...	0.81
A	young	girl	holding	a	...	0.27
...
A	girl	sitting	next	to	...	0.78

Figure 1: An example of evaluating texts by a neural network by a discriminator

From the presented image, you can see that words used several times in different sentences receive a better and more average grade. In the process of generating text, the neural network generator can generate a good beginning of a sentence, but the end of a sentence may be of poor quality. As a result, the discriminator neural network will give a poor estimate of the entire generated sequence. Thus, increasing the number of examples in batch size allows for a better estimate of the generator neural network.

In paper [3], it is proposed to use the Monte Carlo algorithm to assess the quality of the generated samples. This approach allows you to increase the quality of the text evaluation, as well as the use of an increased batch size.

4. Assessment of the quality of text generation according to the BLEU metric

In this work, the text quality is assessed based on the BLEU [9] metric implemented in the nltk library.

Before evaluating sample texts, the placeholder word is removed. This word is used to increase the length of the text to a given length during training. This operation allows you to get a better assessment of texts, since the matches of placeholder words are not taken into account when evaluating samples.

In this work, when evaluating samples, the smoothing method is used. This smoothing method is described in paper [10]. The proposed modification allows you to get a better assessment of the quality of texts.

5. Test results of the considered approaches

The neural network SeqGAN was developed and tested on the task of word-by-word generation of short texts [8]. Implementation [8] and testing of the SeqGAN neural network were carried out on a data sample with captions to images from the COCO Image Captions sample [4].

To assess the quality of text generation according to the BLEU metric, samples from the training sample (written by people) were tested on a test sample. The results obtained are used for subsequent comparison with the generation results. The results of testing 500 random examples from the training sample are presented in Table 1.

Table 1

Test results for examples written by humans

Data sampling	BLEU-2	BLEU-3	BLEU-4	BLEU-5
COCO Image Captions	0.557	0.443	0.319	0.212

The quality of the generation of texts generated by the LSTM neural network trained on the basis of the MLE approach was assessed. The quality was assessed based on the BLEU [10] and negative log-likelihood (NLL) [11] metrics. The test results are presented in Table 2:

Table 2

MLE-trained LSTM test results

Data sampling	BLEU-2	BLEU-3	BLEU-4	BLEU-5	NLL
COCO Image Captions	0.523	0.384	0.248	0.149	1.961

Table 3

SeqGAN test results

Data sampling	Batch size	BLEU-2	BLEU-3	BLEU-4	BLEU-5	NLL
COCO Image Captions	40	0.543	0.412	0.280	0.177	2.006
	400	0.516	0.388	0.265	0.169	1.960
	4000	0.521	0.394	0.270	0.173	1.953

The generator neural network in SeqGAN is an LSTM trained using MLE. Table 3 shows the results of additional LSTM training using the SeqGAN neural network. Lot sizes of 40, 400 and 4000 samples were chosen. This choice is based on the formation of a larger difference in the size of batches, 10 and 100 times.

Examples of texts generated using a neural network trained by SeqGAN are presented in Table 4. For each example, an estimate is given according to the BLEU metric [10].

Table 4

Examples of generated texts

Data sampling	Batch size	№	Examples of texts	BLEU-2	BLEU-3	BLEU-4	BLEU-5
COCO Image Captions	40	1	A baseball player sliding off a base ready to swing .	0.217	0.064	0.035	0.024
		2	A person sitting on top of an orange room , two chairs and two large plants .	0.791	0.624	0.455	0.306
		3	A person is sitting on top of a couch with a bag .	0.640	0.610	0.532	0.432
	400	1	Tennis player players one of the pitcher in an vests opening .	0.251	0.070	0.037	0.026
		2	A populated dog with a side fan painted laptop while sitting on top of it .	0.658	0.577	0.429	0.335
		3	A child is sitting in front of a brick wall of a building .	0.692	0.643	0.575	0.486
	4000	1	A young woman is holding a piece of pizza .	0.486	0.374	0.235	0.114
		2	A person is standing in the middle of a field .	0.538	0.525	0.510	0.463
		3	A person standing on a lush green field next to a large tree .	0.692	0.683	0.641	0.555

Figure 2 shows the change in the quality of the neural network assessment according to the NLL metric during training based on MLE and SeqGAN.

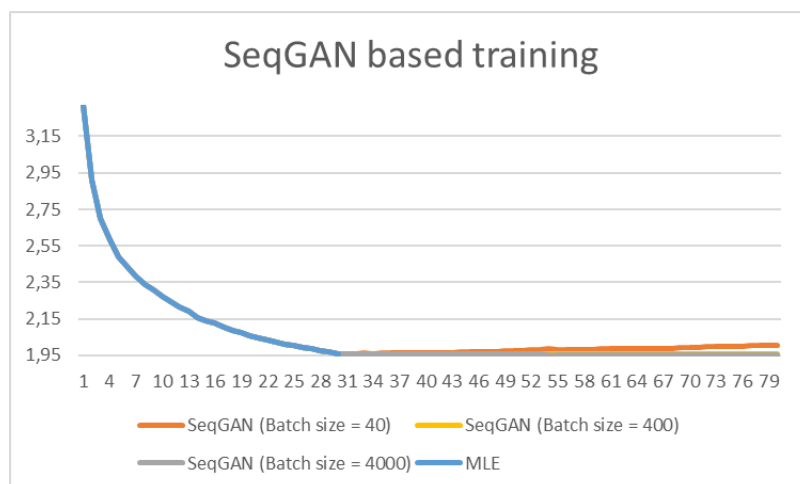


Figure 2: Comparison of batch sizes in the training process based on SeqGAN

Figure 3 shows the process of training a SeqGAN-based neural network using different batch sizes.

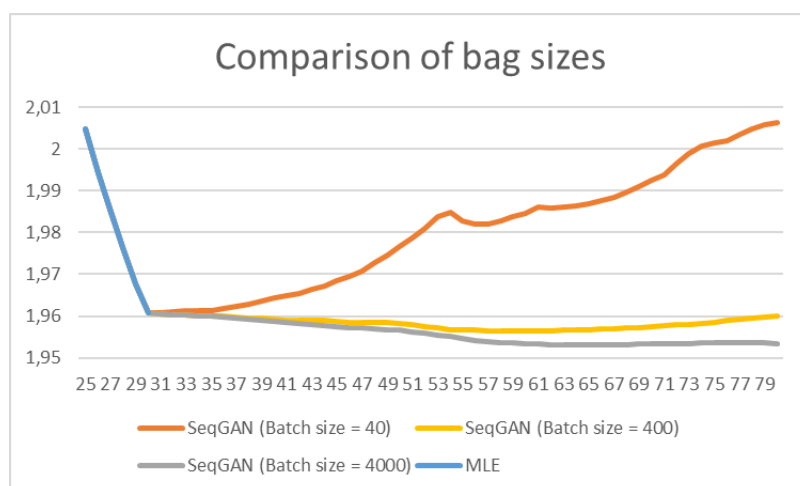


Figure 3: Comparison of the impact of batch size on the learning process with a graph focus on SeqGAN results

Based on the graphs and the results presented in the tables above, we can conclude that the batch size affects the final quality of the neural network training according to the NLL metric. It should also be emphasized that in this work, in the process of training the SeqGAN neural network, the Monte Carlo-based algorithm used in the original work [3] is not used. The Monte Carlo algorithm can improve the quality of training a neural network, but is not considered in this paper.

Using the SeqGAN neural network improves the quality of text generation using the BLEU and NLL metrics. The improvement in text quality can be seen by comparing the learning outcomes based on MLE and SeqGAN. Increasing the batch size reduces the error in the NLL metric. The disadvantage of this approach is the increase in the training time of the algorithm.

When using the batch size parameter equal to 40, the text quality according to the BLEU metric is better than when using the batch size equal to 400 and 4000. An increase in the quality of the text may be associated with a decrease in its diversity, as indicated by a large error in the NLL metric. This can also be evidenced by the higher accuracy of the discriminator, equal to 86%, while when using packets 400 and 4000 it is 81%. It should be noted that the quality of text generation by the neural network trained on the basis of the MLE and SeqGAN approach is inferior to the examples from the training sample according to the BLEU metric.

6. Conclusion

As part of this work, the MLE and SeqGAN approaches were implemented and tested. Based on the presented results, we can conclude that the SeqGAN-based approach allows to increase the quality of text generation according to the NLL and BLEU metrics. In this work, the Monte Carlo algorithm proposed in paper [3] was not used.

The LSTM neural network was trained based on the SeqGAN approach using different batch sizes. Based on the results obtained, it was concluded that the increased batch size makes it possible to increase the training quality of the LSTM neural network.

It should be noted that the quality of text generation by the neural network trained on the basis of the MLE and SeqGAN approach is inferior to the examples from the training sample according to the BLEU metric.

In this work, the software implementation of the investigated approaches in the Python language was carried out. The software implementation is presented on the site [8] and is based on the work [9].

7. Acknowledgements

This research was supported by Tomsk Polytechnic University Competitiveness Enhancement Program.

8. References

- [1] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, *Neural Computation* 9(8) (1997) 1735–1780. doi: 10.1162/neco.1997.9.8.1735.
- [2] J.S. Cramer, *Econometric Applications of Maximum Likelihood Methods*, Cambridge University Press, 1986. doi: 10.1017/CBO9780511572050.
- [3] L. Yu, W. Zhang, J. Wang, Y. Yu, SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient, in: *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 2852–2858. arXiv:1609.05473.
- [4] X. Chen, H. Fang, T.-Y. Lin, R. Vedantam, S. Gupta, P. Dollar, C.L. Zitnick, (2015) Microsoft COCO Captions: Data Collection and Evaluation Server. arXiv:1504.00325.
- [5] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, J. Wang, (2018) Long Text Generation via Adversarial Training with Leaked Information, in: *The Thirty-Two AAAI Conference on Artificial Intelligence*. vol. 32. no. 1. pp. 5141–5148. arXiv:1709.08624.
- [6] Sampling Image COCO, 2020. URL: <https://github.com/CR-Gjx/LeakGAN/tree/master/Image%20COCO/save>
- [7] SeqGAN neural network implementation, 2020. URL: <https://github.com/NikolayKrivosheev/Generation-of-short-texts-SeqGAN>
- [8] SeqGAN, 2020. URL: <https://github.com/suragnair/seqGAN>
- [9] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, BLEU: a Method for Automatic Evaluation of Machine Translation, in: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002, pp. 311–318. doi: 10.3115/1073083.1073135.
- [10] B. Chen, C. Cherry, A Systematic Comparison of Smoothing Techniques for Sentence-Level BLEU, in: *Proceedings of the Ninth Workshop on Statistical Machine Translation*, 2014, pp. 362–367. doi: 10.3115/v1/W14-3346.
- [11] C. Nikolenko, *Deep learning. Immersion in the world of neural net-works*, Piter, 2018. (In Russian). ISBN: 978-5-496-02536-2.