Neural Machine Translation

David Barbas Rebollo

Automatic Translation
January, 2021

1 Introduction

The objective of this laboratory assignment is to use the tool nmt-keras[1], a neural translation toolkit, to train models for automatic translation with pairs of bilingual phrases. The results are then evaluated using the BLEU[2] metric. This is a continuation of our previous lab assignment to compare the results obtained via statistical machine translation.

2 Experiments

The are four different tasks in this lab assignment which involve modifying different parameters and comparing its effect on the performance with respect to the initial model, the one develop during the laboratory class. All experiments used an attention recurrent neural network encoder-decoder, 5 epochs of training and the initial learning rate of 0.001, unless commented differently.

2.1 Task 1

This exercise compares how different values of the word embedding size of the source and target, these two values will be equal, affect its performance.

Table 1: Results comparison of task 1, varying the word embedding size.

Experiment	Embedding Size	Model size	Optimizer	BLEU
Exp1.1	32	64-64	Adam	89.50
Initial	64	64-64	Adam	94.31
Exp1.2	128	64-64	Adam	96.92
Exp1.3	256	64-64	Adam	97.80

As expected, a higher performance is achieved by increasing the complexity of the represenation of the words to be translated, while using a smaller representation hinders the model's performance.

2.2 Task 2

In the second task, the size of the encoder and decoder are varied simultaneously to investigate how simplifying or making more complex the structure of the network affects its performance.

Table 2: Results comparison of task 2, varying the size of the encoder-decoder model.

Experiment	Embedding Size	Model size	Optimizer	BLEU
Exp2.1	64	32-32	Adam	91.47
Initial	64	64-64	Adam	94.31
Exp2.2	64	128-128	Adam	94.19
Exp2.3	64	256-256	Adam	94.73

The results suggest that the performance increases as the complexity of the network increases. However, in this experiment the increase in performance is much lower than in the previous one. Meaning, that there must be overfitting by behalf of the model.

2.3 Task 3

In this task, the objective is to experiment with different learning algorithms and compare their different performance. For this task, the optimizers considered are Adadelta and Adagrad.

Table 3: Results comparison of task 3, varying the optimizer method.

Experiment	Embedding Size	Model size	Optimizer	BLEU
Initial	64	64-64	Adam	94.31
Exp3.1	64	64-64	Adadelta	77.05
Exp3.2	64	64-64	Adagrad	73.94

As expected, Adagrad obtains the worst BLEU value, followed by Adadelta and finally Adam shows the best performance. This correlates with the fact that Adadelta was designed to have a better performance than Adagrad and Adam with respect to Adadelta.

2.4 Task 4

In the last task of the lab assigment, transformers are used instead of attention recurrent neural networks. Also, experimentation is done with different model sizes to have a more well-rounded and wholesome evaluation of this type of model.

Table 4: Results comparison of task 4, varying the type of model and its size.

Experiment	Embedding Size	Model size	Optimizer	BLEU
Initial	64	64-64	Adam	94.31
Exp4.1	32	32	Adam	64.52
Exp4.2	64	64	Adam	87.85
Exp4.3	128	128	Adam	91.76
Exp4.4	256	256	Adam	85.15

Given the obtained results, neither of the experiments done outperforms the initial experiment. Meaning, that attention recurrent neural networks are better for this task.

3 Conclusion

In this lab assigment, there have been serveral experiments conducted with nmt-keras[1] using neural networks. Futhermore, the effect of different hyper-

parameters have been analyzed to describe their effect on a model's performance. Additionally, the experiments were done in the virtual enviroment of the university meaning the were little resources for training. Despite of this, the toolkit is fairly easy to use, making easier the experimentation. Finally, it has been observed how, the word embedding size and the optimization algorithm take a huge role in the performance of a model. This allows to obtain better results than via the use of statistical machine translation methods.

4 Bibliografía

References

[1] Álvaro Peris and Francisco Casacuberta. *NMT-Keras: a Very Flexible Toolkit with a Focus on Interactive NMT and Online Learning.* The Prague Bulletinof Mathematical Linguistics, 111:113–124. 2018.

[2] Papineni, Roukos, Ward, Zhu, W. J. . *BLEU: a method for automatic evaluation of machine translation.* ACL-2002: 40th Annual meeting of the Association for Computational Linguistics. pp. 311–318. 2002.