MISC Research: preventing short greeting and ending pattern

Lisa Hensens

January 4, 2021

Problem

The problem that is occuring with the neural network (NN) could be overfitting, which entails that the neural network is too closely trained on the training data and does not fit unforeseen data and thus may predict this unforeseen data reliably. Our chatbot now predicts short messages, independent from the meaning, as a greeting or ending based on the pattern of the input from the user. The following screenshot captures this problem: Overfitting violates Occam's

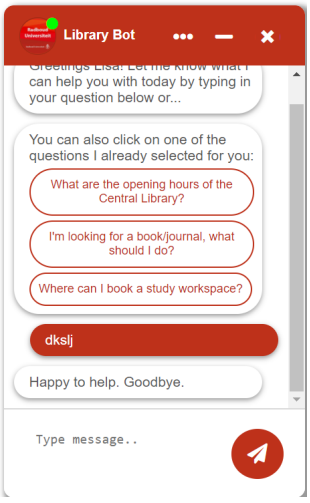

Figure 1: Figure 1: Overfitting example in the chatbot

razor by using a more complicated approach than is optimal, e.g. learning too many parameters.

Solutions

There are a number of solutions that we can try to prevent overfitting in our neural network.

- **Using more data** to train the NN and thus our chatbot. When having a larger dataset for training that contains larger input sequences for greeting and ending tags, the NN could learn better fitting parameters instead of depending on the short pattern for the greeting and ending tags.
- **Early stopping** by inspecting the prediction loss to decide when to stop reducing loss on the training data. But this approach in general is not enough to prevent overfitting.
- **Dropout** refers to randomly omitting units during the training process [1]. This prevents the units to become too dependent on each other. Because of this, the units will learn more robust features. Dropout can be seen as a form of ensemble learning, because it regularizes the weigths over units and would thus work for smaller neural networks such as our network for the chatbot.
- **Cross-validation** or **k-fold cross-validation** for the training process. This will derive more accurate values for the parameters when training the NN.
- **Remove irrelevant input features** [2].There are several feature selection heuristic approaches to spot the irrelavant features and remove them.

Conclusion

The most likely issue that occurs in our NN is overfitting. We could solve this by simply training on more data that contains more examples per tag and thus learning more complete parameters that can fit unforeseen data better than it does now. Because we do not have a lot of data now or can get a large amount of data quickly, we can use another approach on preventing overfitting by using e.g. cross-validation or dropout.

References

[1] `https://en.wikipedia.org/wiki/Dilution_(neural_networks)#cite_note-MyUser_Jmlr.org_July_26_2015c-3`

[2] `https://elitedatascience.com/overfitting-in-machine-learning#how-to-prevent`