MIE324 Assignment 4

Daniil Orekhov - 1002302938

November 2018

3 Preparing the Data

3.1

Number of ones in train: 3200
Number of zeroes in train: 3200
Number of ones in validation: 800
Number of zeroes in validation: 800
Number of ones in test: 1000
Number of zeroes in test: 1000

7 Grading Experimental and Conceptual Questions

7.1

Model	Train		Validation		Test	
	Error	Loss	Error	Loss	Error	Loss
Baseline	0.0016	0.0227	0.0863	0.2324	0.086	0.2295
RNN	0.0	0.0001	0.1194	0.7820	0.124	0.8290
CNN	0.0	0.0006	0.1156	0.3790	0.114	0.4020

Table 1: Training values for the three different types of models

These values suggest that baseline model performs the best. There is also almost no difference between test and validation, since we did not do any hyperparameter search, so we never made the models overfit to hyperparameters.

7.2

The baseline model is able to interpret general idea of the sentence, but it does not analyze the specific word choice and order. Thus, it is very difficult to even over-fit the baseline model, since it is essentially impossible to learn the concept of "objectivity" solely looking at the the topic of the sentence.

7.3

Model	Train		Validation		Test	
	Error	Loss	Error	Loss	Error	Loss
Default	0.0	0.0001	0.1194	0.7820	0.124	0.8290
W/o using pack_padded_sequence	0.0	0.0001	0.1231	0.8691	0.1415	0.9812
Using Iterator	0.0	9.334e-05	0.1188	0.4546	0.1175	0.4978

Table 2: Training values for RNN, examining the effect of pack_padded_sequence and Iterator

It seems like padding doesn't really affect the results by much, so it is hard to examine the effect of this change. However, one could expect that the effect of this change would be minimal, since for most sentences the order of words might not matter that much.

7.4

The kernels in the CNN represent the idea of phrases. The kernels "learn" to assess each combination of neighbouring words and evaluate the subjectivity of it. Thus, CNN doesn't really understand what the sentence is about, it can only assess the phrases used.

7.5

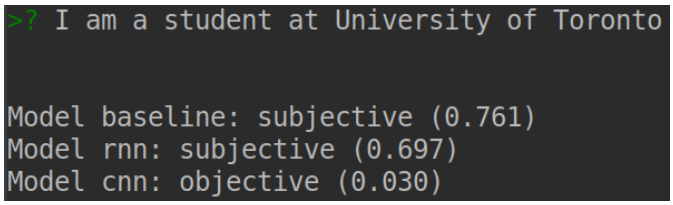

Figure 1: Examples of models performance on objective statement

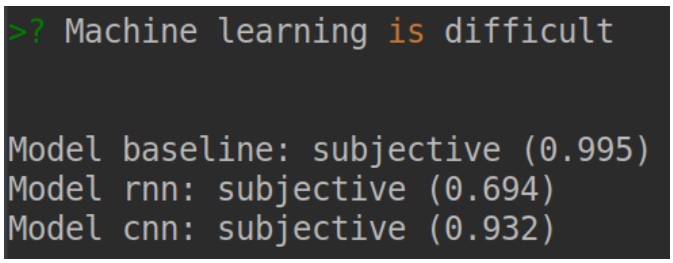

Figure 2: Examples of models performance on subjective statement

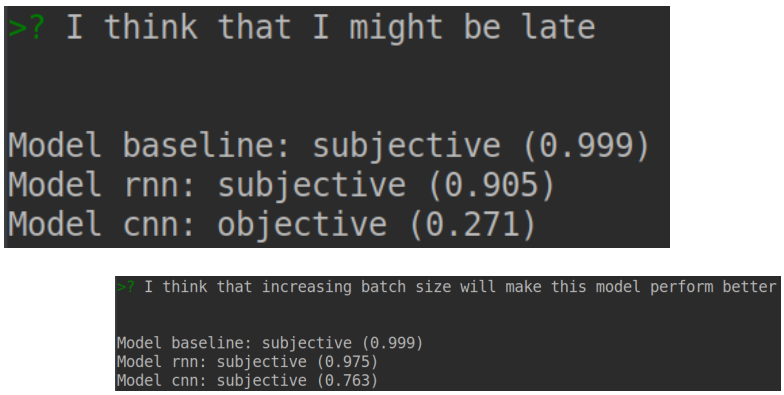

Figure 3: Examples of models performance on borderline statements

It is not obvious how to evaluate the model performances based on how they assess individual sentences, since the performance varies greatly and it is not very representative of the models abilities. In the examples in the figures, the baseline is performs great at evaluating the subjective sentence, but it is also 100% confident in the borderline cases; RNN treats a subjective and an objective sentence equally and is very confident about borderline sentences as well; CNN seems to be performing the best with the given examples, correctly identifying the objective and subjective sentences and not being over-confident in borderline cases. However, CNN performing the best might suggest that it is just the phrases I have been using indicate how subjective each sentence is.

7.6

7.6.1

I spent about 8 hours on this assignment.

7.6.2

Understanding the assignment descriptions was very challenging. The process of preparing data was slightly difficult, but once my peers explained what I was supposed to do,it was quite easy.

7.6.3

The fact that we were doing NLP in general is fun, but I think it would be nicer to have better (easier) topic for this assignment, that would yield more accurate results.

7.6.4

Both the assignment description and the API were very confusing, so it was a challenge to actually understand the expectations for the assignment and how to implement them.

7.6.5

Some of the guides linked in the assignment description were quite helpful.