Introduction

The objective of classification in machine learning and natural language processing is to designate one or more labels to a given input based on its characteristics. In this report, we present the outcomes of a classification task performed with a PyTorch LSTM model trained on a text data set. The training set consisted of sentences annotated with different part-of-speech tags, and the model was trained to predict these tags when given a sentence as input. The dataset was partitioned into training and test sets, and the model was evaluated based on its ability to accurately predict part-of-speech marks on the test set. The classification report provides a summary of the model's performance in terms of precision, recall, and accuracy for each label, as well as macro- and weighted-averaged scores. The results provide insight into the model's strengths and deficiencies and can influence future performance enhancements.

Results

The table below displays the outcomes of a classification task utilising a PyTorch LSTM model trained on a set of text data annotated with various part-of-speech identifiers. The model's ability to accurately predict part-of-speech tags on a held-out test set was evaluated, and the results are summarised in terms of precision, recall, and accuracy for each label, as well as macro- and weighted-averaged scores. The table provides a comprehensive overview of the model's performance on each label and can be used to identify areas in which the model performs well and those in which it requires further refinement.

Class	Precision	Recall	F1-Score	Support
<bot>	1.00	1.00	1.00	586
<eot>	1.00	1.00	1.00	586
adj	0.93	0.96	0.94	220
adp	0.99	1.00	1.00	1434
adv	0.95	0.78	0.86	76
aux	0.97	0.99	0.98	256
cconj	1.00	1.00	1.00	109
det	0.99	0.87	0.93	512
intj	1.00	1.00	1.00	36
noun	0.99	0.99	0.99	1166
num	0.97	0.88	0.92	127
part	0.96	0.98	0.97	56
pron	0.86	0.98	0.92	392
propn	0.98	1.00	0.99	1567
verb	0.99	0.98	0.98	629
Accuracy			0.98	7752
Macro Avg	0.97	0.96	0.96	7752
Weighted Avg	0.98	0.98	0.98	7752

Table 1: Classification report

Inferences

Overall, the model's accuracy on the majority of classifications is 0.98, which is excellent. The model's precision, recall, and F1-score are all greater than 0.90 for the adp, aux, cconj, det, intj, noun, part, pron, and propn labels. The model's performance on the adj, adv, and num labels is marginally inferior, with precision, recall, and F1-score ranging from 0.86 to 0.94. The model's low recall for the adv and num labels suggests that it may struggle to correctly predict these labels for certain inputs. The model's low precision for the pron label suggests that it may produce a greater number of false positive predictions for this label than for others. Both the macro- and weighted-averaged scores are greater than 0.96, indicating that the model performs well overall. The weighted-average score is marginally higher than the macro-average score, indicating that the model performs better with labels that occur more frequently.

Hyperparameters

The input size and hidden size were both set to 300, indicating that the model accepts 300-length input sequences and has a 300-size hidden state. The number of LSTM layers has been set to 1, so the model contains a single LSTM layer. The vocabulary size was set to the length of the vocabulary in the training data, which corresponds to the number of unique words in the training data. The output size was set to the length of the tag vocabulary in the training data, which corresponds to the number of unique part-of-speech tags in the training data.