

Anton Goretsky

1) Accuracy

Running the model with the default parameters 5 times, by the end of 20 epochs, the test set accuracy averaged around 84.25%.

2) Dev Set Misclassification

Some examples that could be misclassified include: "Name this document which after being implemented led to a schism with Prussia allowing for a second Partition of Poland", which is labeled as "history" but can be confused with literature, as the question specifies a document.

Another example is a question like the following: "This historian , who wrote the massive three-volume tome The Civil War : A Narrative , is perhaps best known for providing much of the commentary in Ken Burn 's documentary The Civil War ." This is labeled as "history", meanwhile much of the text concerns a literary text, albeit one about history.

If we logically consider the categories of Literature and History, they are inherently related in many cases. For example, if we consider a question regarding historical non-fiction writing, then the category could easily be misinterpreted, given many words relating to historical events or the past. Inherently there may be historical events in literature, and literary elements in history.

I would argue that the science class would be easiest to classify, given that its text examples tend to have more separately themed, technical language, unless once again we discuss a scientific achievement in history, or a book about scientific achievements, and then we fall into a gray area.