W266 Sp'22

Authorship Attribution: HuggingFace Keras Model with BERT Tokenization

Tessa de Vries, Shanie Hsieh, & Shirley Jiang

Problem Statement

- Author attribution/identification is increasingly more difficult to solve
  - Growing number of texts and authors
  - Plagiarism
  - Anonymous authorship
- Build a model that can accurately predict the author of a text
  - Learn writing style
  - Project Gutenberg

Motivations

- Lots of questions:
  - Aside from the character names, location/setting, what else about a text can identify an author? People would usually identify a piece of text with those features, but a deep learning model will be able to see deeper.
  - Early forms of authorship attribution were limited in scope, with a few labels and labeled texts. Our issue is an infinite pool of authors. Anyone who can write can be an author, so how do we deal with many labels?
- Want to use BERT language models to explore textual data
- Want to see how the model interprets texts
  - Does the model confuse an author for another? What about the author's work and the author's background might confuse the model?

Methods

- Final form: paragraph level segmentation → sampled 150 paragraphs per author
  - 20 total authors
- Hyper tuned Keras model
  - Optimizers: Adam and SGD
  - Epochs: 3, 5, 10
  - Learning Rate: 1e-5, 3e-5, 5e-5, 1e-2, 5e-2
  - Decay Steps: 10,000 and 50,000
  - Decay Rate: 0.7 and 0.9

Results

We saw the model evolve in many ways!
The following is a sampling of our model prediction outcomes.

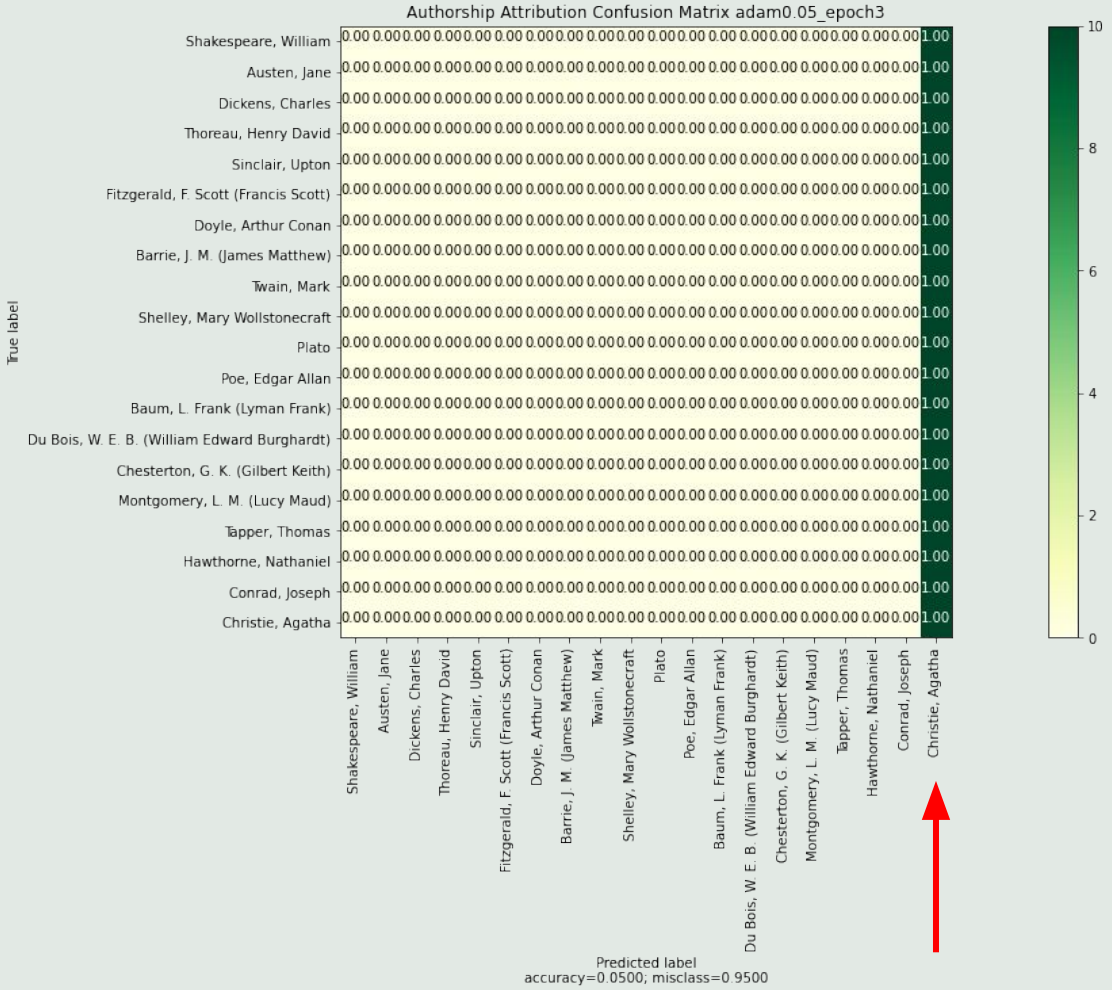
Adam Optimizer

Optimizer	Learning Rate	Epochs	Validation Accuracy
Adam	5e-5	3	0.6450
Adam	5e-5	5	0.6917
Adam	5e-5	10	0.6217

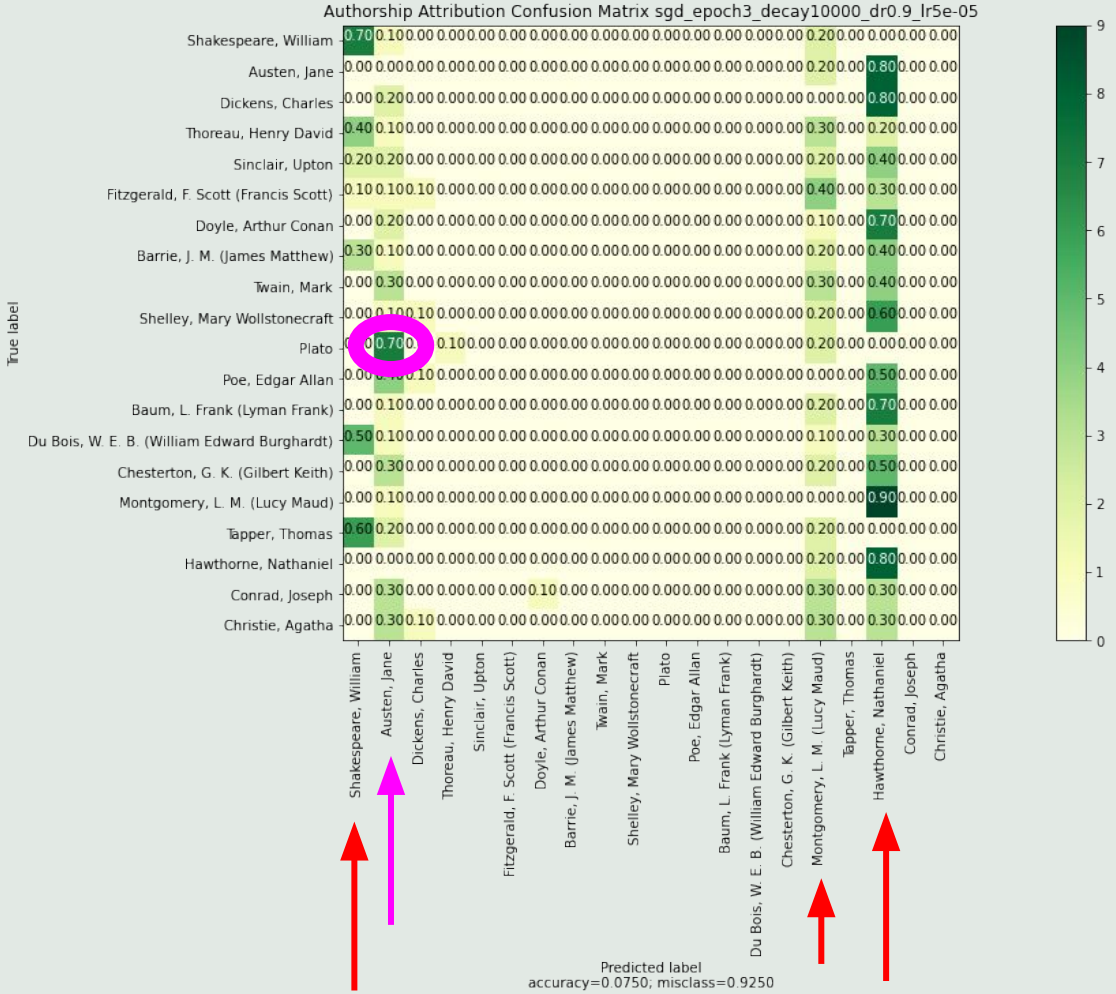
SGD

Optimizer	Learning Rate	Decay Steps	Decay Rate	Epochs	Validation Accuracy
SGD	5e-5	10000	0.9	3	0.5417
SGD	1e-2	10000	0.9	3	0.5200
SGD	1e-2	10000	0.9	5	0.0883

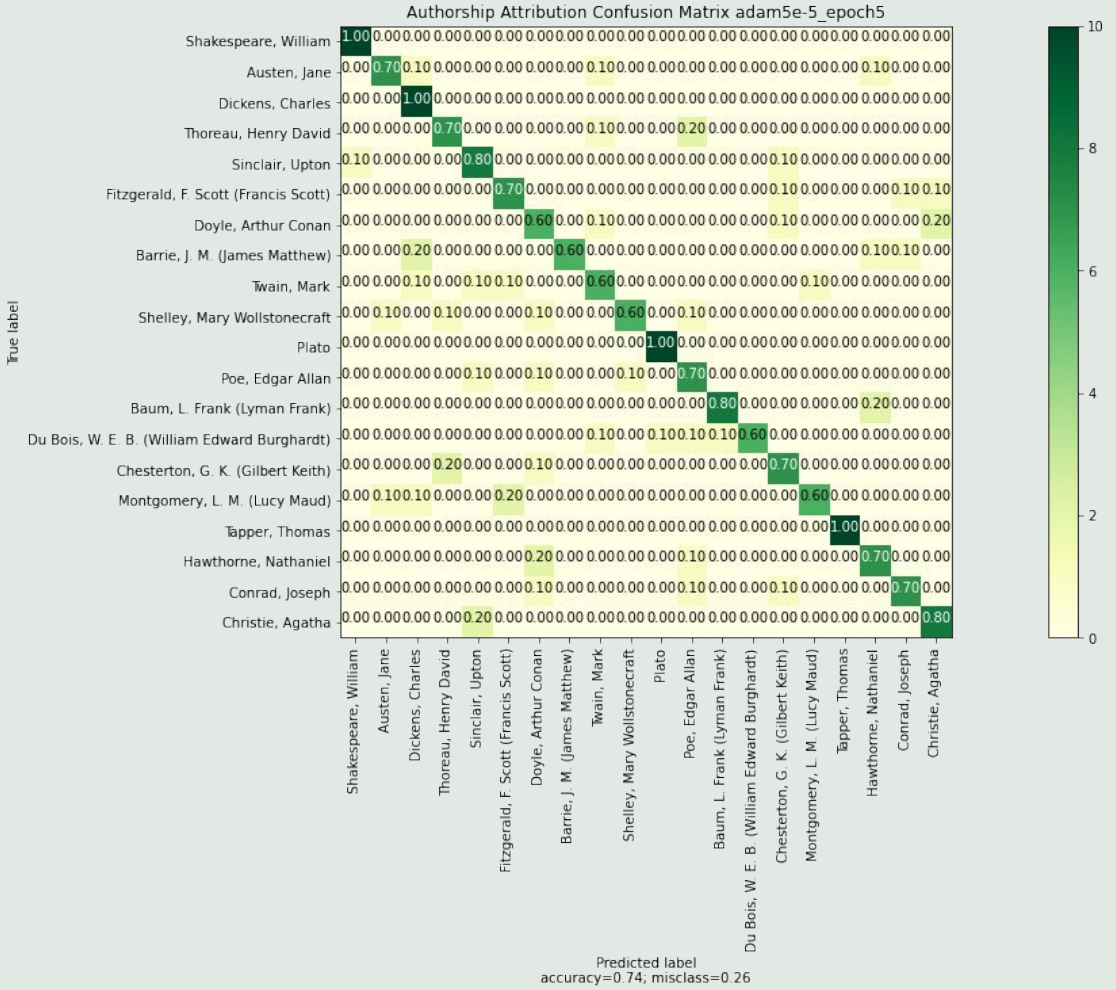
Predicting only one author



Guessing mostly the same label, but not quite



Best Model Accuracy



Conclusion

- Limitation: Larger GPU and RAM needed
- Future work may explore
  - Adding more authors of varied backgrounds and more examples of text
    - Getting more compute power to allow for this larger dataset processing
  - Using our model on other datasets
  - Exploring BERT models, exploring other pre-trained BERT tokenizers (beyond bert-based-case)
  - Slackening BERT token limits with BigBird to enable document-level authorship attribution as opposed to our paragraph-level authorship attribution

Thanks!