Author Verification: A Study of two Methods

Objective for both methods: To create a model for author verification on a pair of texts drawn from emails, news articles and blog posts. Author verification is a binary pairwise classification task where the label of 1 indicates the texts aere written by the same author and a label of 0 indicates the texts are written by different authors.

Feature Extraction with Multi-layer Perceptron(MLP)

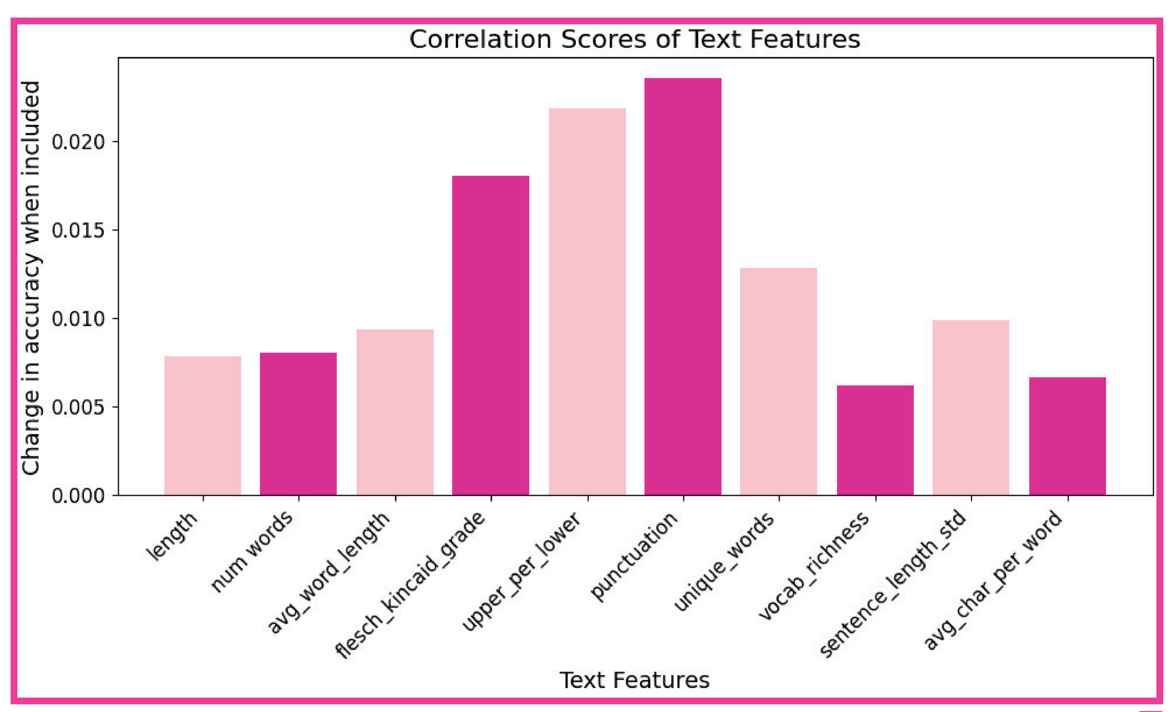
01.Introduction and Background

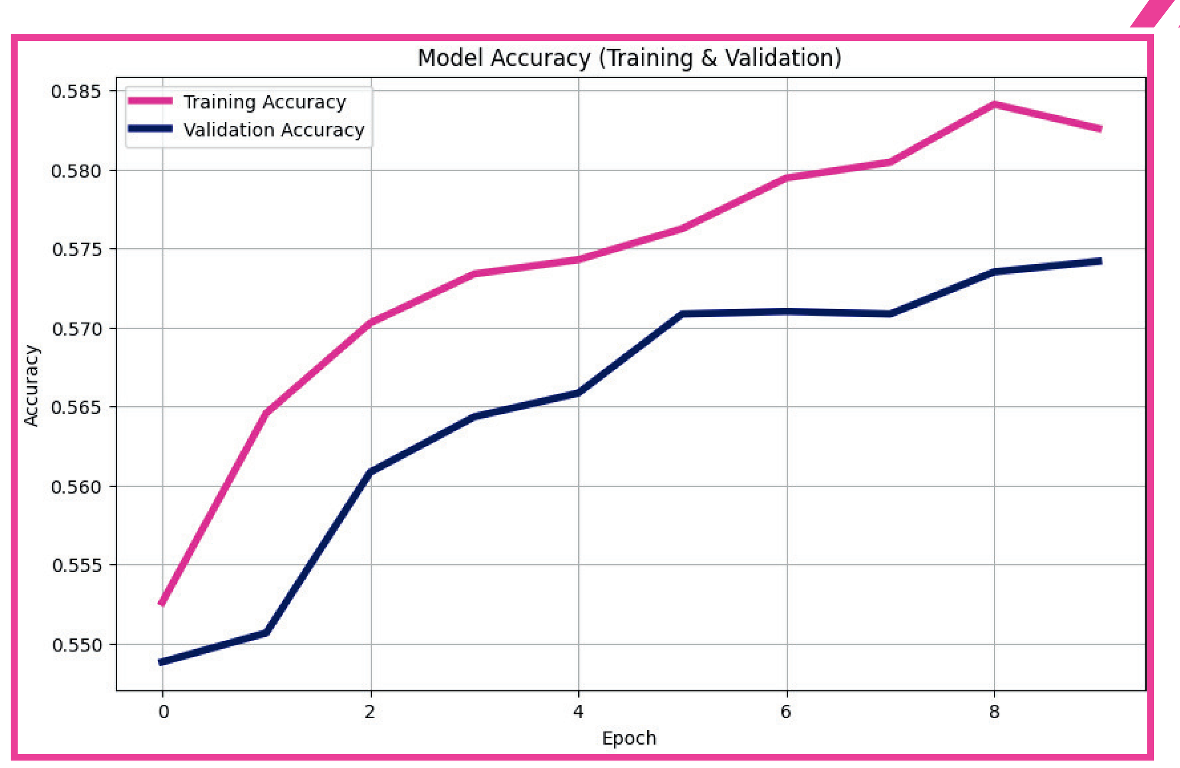
Feature extraction is one of the main approaches used for author verification.This model is initially based on **GLAD: Groningen Lightweight Authorship Detection** developed by Manuela Hürlimann et al in 2015. Many of the features used in this model are based off the features in GLAD such as **punctuation and letter casing**. Other features such as entropy and compression were attempted but did not significantly help. Our **implementation deviates from GLAD** as seen by the other features below.

02. Methodology

Craft **10 handmade features** that make comparisons between the texts from a character to sentence level. These include comparisons in the length of text, **average word length, number of words, readability, capitalisation, punctuation, vocabulary richness and standard deviation in sentence length**. A vector of these scores were then fed into a **MLP with two hidden layers of size 64 and 32** before a single neuron output layer with **sigmoid activation** for binary classification.

03. Experimentation





04. Results

To find the best combination of features, multiple trial training experiments were run. The first graph shows a subset of the results of this trial. The value of each feature is the numerical increase in accuracy when this feature was included. This then shows the features that contribute most significantly to the classification. These **most significant features are punctuation, the ratio between upper and lower case characters, and flesch kincaid grade**. The flesch kindcaid grade is a readability score. The second graph shows the training and validation accuracy during one of these tests. The line graphs illustrate how **both accuracies stop increasing at 10 epochs**, hence the tests were trained on 10 epochs.

05. Conclusion

Overall, this model was successful. By achieving a **top accuracy of 58%**, it **outperformed the baseline method by almost 5%**.

Deep Learning with Bidirectional Gated Recurrent Unit(Bi-GRU)

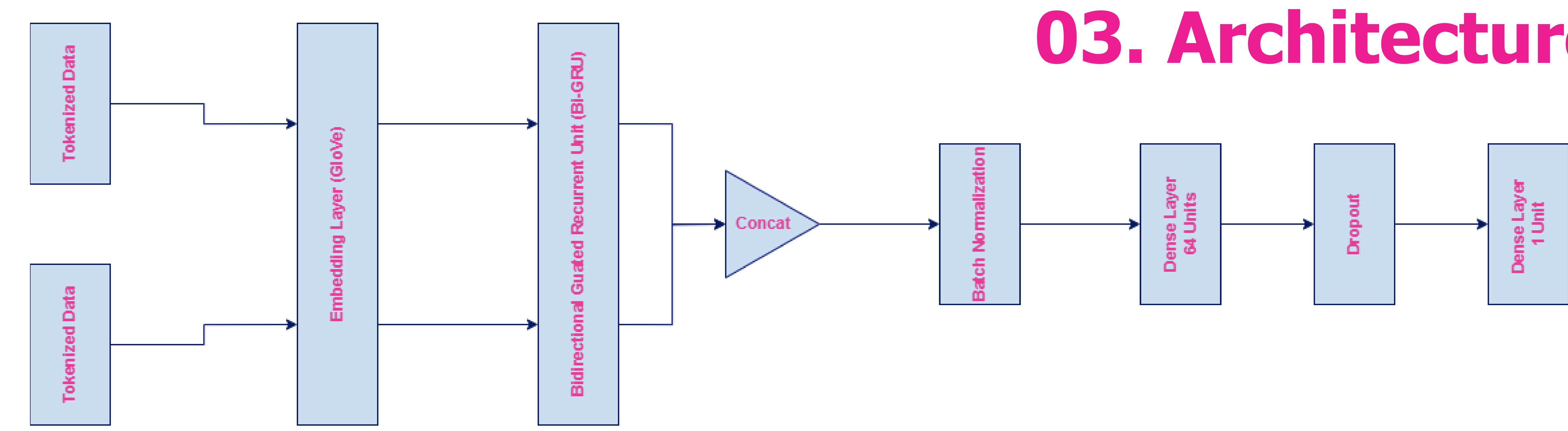
01.Introduction and Background

Using Recurrent Neural Network (RNN) is another method to extract information about the writing style for author verification. **A Siamese Bi-GRU is** described in "Deep Learning based Authorship Identification" by Chen Qian, Tianchang He, and Rao Zhang. The Siamese Bi-GRU architecture described here is implented in the model.

02. Methodology

This model is **siamese Bi-GRU** model. Its architecture consists of a shared embedding layer, followed by a shared Bi-GRU. The weights of the embedding layer are set to **pre-trained Word2Vec** values**.** The outputs of the Bi-GRU are then concatenated together and fed into a **Batch Normalisation layer** and then a **dense layer of 64 neurons**. Finally there is a single neuron output layer with a **sigmoid activation function** for binary classification. To combat overfitting **L2 normalisation** is added to the dense layer and a **dropout layer** is included before the output layer.

03. Architecture



04. Experimentation

Method	Val Accuracy
Dropout Only	0.5038
Adding Regularization and Batch Normalization	0.5189
Adding Word2Vec	0.5440

05. Results

The results of the experimentation show a small general increase when adding dropout, batch normalization, and regularization to combat overfitting. **The addition of pretrained embeddings caused a larger increase in accuracy.**

06. Conclusion

Overall, this model was also successful**.** Through achieving a **top accuracy of 54%, it beat the baseline score by over 4%.**

By Luke Sanderson and Sarah Saad, special thanks to our Google Colab pet