Using Neural Networks to Predict Who Wrote the Disputed Federalist Papers

Brandt Smith

Rollins College

Introduction

The *Federalist Papers*, written by Alexander Hamilton, James Madison and John Jay, are a collection of papers written to convince the people of New York to ratify the new United States constitution. All these documents were written under the same pseudonym "Publius" which left their real author unknown. Before Alexander Hamilton died, he claimed that he wrote 63 of the 85 papers. After his presidency, James Madison claimed that he wrote several papers that overlapped with Hamilton's list. Due to this overlap, I am unsure of who wrote the 12 disputed *Federalist Papers*.

In this paper I discuss a method for determining the author of the 12 disputed *Federalist Papers*. To do this, I utilize a neural network to analyze the differences in languages used amongst the known *Federalist Papers* and create a model that I use to predict the author of the disputed papers. Throughout the remainder of this paper I will discuss:

- The Related Work section discusses previous work on computational algorithms and their use in determining who wrote the disputed *Federalist Papers*.
- Our Methods section describes our data set and our choice of Neural Network model for predicting the author of the papers.
- Results section summarizes the results I obtained from our model as well as how changes in the parameters of our model effect its output.

Related Work

Previous research on determining the author or authors of the disputed *Federalist Papers* often uses the statistical analysis of the language used in the non-disputed papers. This statistical analysis is then used to analyze the language used in the disputed papers and match them to the author that used similar language in their known papers.

(Tweedie, F.J., Singh, S. & Holmes, D.I., 1996) uses stylometry, a statistical analysis of literacy style, to predict the author of the disputed *Federalist Papers*. A wide range of literacy styles were considered when defining what the best way to distinguish the differences between the authors. These include sentence lengths, number of nouns, articles or pronouns, counts of the most common words in the text and words that only occur once in the text. Because stylometry is essentially a pattern recognition problem, the authors created a neural network they used to model the patterns in the text that consisted of an input, hidden and output layer of nodes. Using their neural network, the authors predicted that Madison wrote the disputed papers which is consistent with other work performed.

(Bosch, R. A., & Smith, J. A., 1998) associate each of the *Federalist Papers*, including the disputed papers, with a point in a 70-dimensional space. The 70 dimensions are represented by the number of times per 1000, that each of the 70 "function words" appear in the paper. The authors create a hyperplane that separates the known papers by their authors based on their unique points in the 70-dimensional space.

Once the authors of this paper have their hyperplane that separates the papers into two distinct groups based on their author, they can then create 70-dimensional points for the disputed papers. Those points can be assigned to the author that aligns with the side of the hyperplane that this point belongs too. The created hyperplane successfully predicts that the disputed *Federalist Papers* were written by Madison.

(Fung, G., 2003) use a similar method of generating a hyperplane as (Bosch, R. A., & Smith, J. A., 1998). However, this hyperplane is only in three dimensions. Fung took the same idea and by performing a "validation procedure," limited the number of parameters from 70 to the 3 that best split up the papers by their respective authors. This limited the number of linear programs solved from 57225 in (Bosch, R. A., & Smith, J. A., 1998) down to 65 which includes the process involved in the tuning and validation phase. The hyperplane generated predicts that all 12 disputed papers belong to Madison however, the main contribution in this paper is that it requires significantly less calculation to create the hyperplane.

(Tearle M., Taylor K., Demuth H., 2008) utilize an artificial neural network to generalize the known *Federalist Papers* into a model that allows the neural network to predict the author of papers that is unknown. Thirty-two metrics used by the authors to separate these papers by author include average sentence length, possessive apostrophes per sentence, quotation marks per sentence, dashes per sentence, and more. The authors used their algorithm to predict the differences between works written by Shakespeare and Marlowe as well as the *Federalist Papers*. This algorithm predicted that all the disputed papers were written by Madison as well.

Methods

Data

The data I used for our *Federalist Papers* is identical to the data used in (Bosch, R. A., & Smith, J. A., 1998). The data consists of 118 pairs of lines of text, each pair of lines represents one of the *Federalist Papers*. The first line in the pair holds two numbers: the number of the Federalist Paper and a number that specifies who the author is (1 for Hamilton, 2 for Madison and 3 for undisputed). The second line of the pair contains 70 decimal numbers that represent the number of times per 1000 words that the 70 functional words (Figure 1) exist in the paper. From this data, I separate the papers into three categories: Hamilton's papers, Madison's paper and the disputed papers.

Figure 1: 70 Functional Words

1	a	15	do	29	is	43	or	57	this
2	all	16	down	30	it	44	our	58	to
3	also	17	even	31	its	45	shall	59	up
4	an	18	every	32	may	46	should	60	upon
5	and	19	for	33	more	47	some	61	was
6	any	20	from	34	must	48	some	62	were
7	are	21	had	35	my	49	such	63	what
8	as	22	has	36	no	50	than	64	when
9	at	23	have	37	not	51	that	65	which
10	be	24	her	38	now	52	the	66	who
11	been	25	his	39	of	53	their	67	will
12	but	26	if	40	on	54	then	68	with
13	by	27	into	41	one	55	there	69	would
14	can	28	into	42	only	56	things	70	your

Model

To predict who wrote the disputed *Federalist Papers* I use a neural network. Using Python packages for TensorFlow and Keras, I created a four-layer dense neural network to model our data. Because our model is dense, each node is connected to every node in the layer directly beyond it and so on for future layers. The input layer of our neural network has 70 nodes, one node for each of the 70 function words in Figure 1. Then comes two hidden layers both having 128 nodes and using a sigmoid function as their activation function. The sigmoid function "smushes" the outputs of the hidden layers into values between 0 and 1.

$$\text{Sigmoid Function f(x)} = \frac{1}{1 + e^{-x}}$$

Lastly, our output layer has two nodes, one for Hamilton's papers and the other for Madison's papers. This last layer uses a SoftMax activation function, which normalizes the outputs of this layer into a probability distribution selecting the highest one to be the predicted output.

Upon compiling our model, I use Adam as our optimization function. Adam is a straight forward, computationally and memory efficient gradient-based optimization function for stochastic objective functions. The loss function I used is Sparse Categorical Cross Entropy. I chose this loss function because each of our data points belongs to exactly one class.

Results

Upon using all our testing data, 106 of the 118 pairs of data from (Bosch, R. A., & Smith, J. A., 1998) with our neural network as mentioned above, after 30 epochs our model predicts that 1 of the disputed papers was written by Hamilton and the other 11 paper were written by Madison. These results are very close to the other studies in the related works section. One thought about these results are, what would happen if I change how many epochs, I put the model through or what would happen if I change the model itself.

Let's first look at how changing the number of epochs changes the prediction of who wrote the papers. By changing the number of epochs to 15, the model now predicts that 9 of the 12 papers were written by Hamilton and only 3 of the 13 were written by Madison. This is interesting because I know from previous research that this should not be the case, most of the papers should be predicted to be written by Madison. However, if I raise the number of epochs to 45, the prediction is that 2 of the 12 papers were written by Hamilton and 10 of the 12 were written by Madison. This is closer to what the research suggests, however, this may suggest that there are unique elements of both writers in some papers that may suggest they collaborated on some of the papers.

Lets now look how changing the model itself effects the prediction of the authors after 30 epochs. First, I will remove one of our 128 node hidden layers. By doing this the network predicts 2 of the papers were written by Hamilton and 10 by Madison. If instead I add an additional layer, the network predicts that 1 of the papers was written by Hamilton and 11 were written by Madison. This result is the exact result I received with our original network; however, it took longer to compute due to the additional hidden layer. Lastly, if I create 6 hidden layers, all with 128 nodes, and execute our program, the network again predicts 2 papers were written by Hamilton and 10 were written by Madison. This goes to show that more is not always better.

Finally, I will test our original network with the two hidden layers with different numbers of nodes to see how that effects our results. First, I set them to have 64 nodes each instead of 128 each and that gave us the results of 2 papers for Hamilton and 10 for Madison. Now if I give them 256 nodes instead of 128, I receive results suggesting that Hamilton wrote 1 paper and Madison wrote 11. These results suggest that the more I add to both the more these predictions settle. Lastly, I will give the first hidden layer 128 nodes and second hidden layer 256 nodes. The results of this setup also predicted that Hamilton wrote 1 and Madison wrote the other 11 papers.

Upon making these changes to the neural network, I couldn't get the network to predict that Madison wrote all 12 papers. The biggest change I saw was due to the number of epochs being changed which could be attributed to the network not having enough data/time to be able to set the proper weights for the functional words. Potentially in the future finding the optimal settings that would allow the network to predict all 12 papers were written by Madison would involve changing epochs as well as number of hidden layers and nodes per hidden layer.

Conclusion

In this paper I discuss my implementation of a neural network to predict who wrote the disputed *Federalist Papers*. Using data from (Bosch, R. A., & Smith, J. A., 1998) to model the paper our network was able to predict that Hamilton wrote 1 paper and Madison wrote the other 11. Upon testing I couldn't determine if Madison wrote all 12 papers, however, upon future changes could potentially configure the network in such a way that it can make that prediction.

References

Bosch, R. A., & Smith, J. A. (1998). Separating hyperplanes and the authorship of the disputed federalist papers. The American mathematical monthly, 105(7), 601-608.

Fung, G. (2003, October). The disputed Federalist Papers: SVM feature selection via concave minimization. In Proceedings of the 2003 Conference on Diversity in Computing (pp. 42-46). ACM.

Matt Tearle, Kye Taylor, Howard Demuth, An algorithm for automated authorship attribution using neural networks, *Literary and Linguistic Computing*, Volume 23, Issue 4, December 2008, Pages 425–442, https://doi.org/10.1093/llc/fqn022.

Tweedie, F.J., Singh, S. & Holmes, D.I. Computers and the Humanities (1996) 30: 1. https://doi.org/10.1007/BF00054024.