Finding Publius:
Determining the Authorship of the Federalist Papers through Statistical Analysis

Erick Ortiz, Eduardo Aceituno, Alexis Darnat, Daniel Son
CECS 429/529
December 22, 2018

I. Introduction

The Federalist Papers, a series of eighty-five essays, published in 1787 through 1788, were penned under the pseudonym Publius. The primary use of a pseudonym served as a form of protection from retaliation. This corpus of documents was authored with the intent of overriding the *Articles of Confederation* and ratifying the *Constitution of the United States*. Publius, by historical consensus, was identified as Alexander Hamilton, John Jay, and James Madison. Upon the death of Alexander Hamilton, a list emerged claiming that he alone had authored nearly two thirds of the essays. (Adair 1944) James Madison, too, asserted that a subset of the essays, claimed by Hamilton, were drafted by his pen. A total of eleven essays were left in contention. With further investigation, historian Douglass Adair hypothesized the following authorship: (Adair 1944)

Alexander Hamilton	{1, 6–9, 11–13, 15–17, 21–36, 59–61, 65–85}
James Madison	{10, 14, 18–20, 37–58, 62–63}
John Jay	{2-5, 64}

This was later corroborated by computational analysis of the text in 1964. The contested documents are 49-57, 62, 63. No one truly knows who wrote the essays with absolute certainty. However, by leveraging advances in technology and modern computational power we can conduct statistical analysis of the corpus to provide a solid guess of who wrote the contested documents. Given a training set of labeled documents, the use of statistical text classification and vector space classification provides us with ways of classifying documents. We implemented two classification techniques, Rocchio and Bayesian.

II. Foundations of Classification: Vector Space Model

The classification methods used to determine the authorship of *The Federalist Papers* are rooted in machine learning. “In machine learning, the set of rules or, more generally, the decision criterion of the topic classifier is learned automatically from training data." (Christopher D. Manning 2009) This strategy of classification is also referred to as statistical text classification. Of course, the training data used to determine authorship will derive from the documents that are not in contention and whose authorship is agreed upon by historians.

The Rocchio classification is a "vector space classification that divides the vector space into regions centered around a centroid, one for each class." (Christopher D. Manning 2009) The underlying premise of using vector space classification is given by the Contiguity Hypothesis: "Documents belonging to the same class form a contiguous region and do not overlap with other regions." (Christopher D. Manning 2009). If this holds, documents can be classified into the region they fall into.

Naive Bayes text classification is a supervised probabilistic learning method. The text classification is based on the Bayes theorem which also has the assumption that there is independence within the features. The Naive Bayes text classification relies on a discriminating vocabulary set that represents the document. The problem with the Naive Bayes text classification is that there can be zero frequency for features which makes the probability zero. Laplace Smoothing is applied to correct Naive Bayes text classification.

III. Rocchio Classification

The first step in implementing Rocchio classification is identifying the centroids for each class. To calculate each centroid vector, we need to ensure each document *d* belonging to a particular class *c* is weighted and normalized accordingly. We use the $w_{d,t} = 1 + \ln (tf_{t,d})$ value to weight each term occurring in the document, where $tf_{t,d}$ refers to the number of times a term appears in a document. Additionally, each $w_{d,t}$ is normalized by the Euclidean length L_d, given by:

$$L_d = \sqrt{\sum_t \left(\mathrm{w}_{d,t}\right)^2}$$

Once the vectors for each document were calculated, the centroid was computed by adding all vectors in the class and dividing by the number of documents in the class D_c, as shown in the following formula:

$$\vec{\mu}(c) = \frac{1}{|D_c|} \sum_{d \in D_c} \vec{v}(d)$$

After identifying each centroid and loading the training set, what remains is generating vectors for each disputed document and classifying them. We performed the same steps to create a vector for each disputed document - ensuring all values are weighted and normalized - and compared the vector to each centroid. The comparison is done by calculating the Euclidean distance to each centroid:

$$|\vec{\mu}(c) - \vec{v}(d)|$$

Once we find the closest centroid, we classify the document as a member of the given class.

IV. Naive Bayes Classification

The foundations of Bayesian classification is rooted in probability theory. In order to train the data set to correctly classify the disputed papers, each training set needed to output a discriminating vocabulary set. To accomplish creating the discriminating vocabulary set, we used mutual information for the selection. For each class in the training set, we calculated the I(t,c) for each term within a class. Once we had a score for each term in all the classes, we put the term with their score into a priority queue which enabled us to find the top I(t,c) scores. With having the top scores, we can now choose the size of the discriminating vocabulary set. The discriminating vocabulary set was used to train the Bayesian classifier by calculating

P(t|Hamilton), P(t|Madison), and P(t|Jay) with Laplace Smoothing. The values for each term is essential for scoring the disputed papers. Once the Bayesian classifier is trained, we went through the disputed documents and calculated the score for each class using the formula:

$$c_d = argmax_{c \in C} \left[log(p(c)) + \sum_{t_i \in T_d} log(p(t_i|c)) \right]$$

Since p(c) and p(ti|c) gives us a decimal value, the formula will output a negative number. Getting the max of the arguments, we get prediction of which person wrote that specific document.

V. Results

Our implementation of Bayesian classification did not produce the expected outcome. 50 terms were inserted into the discriminating vocabulary set and we set the vocabulary set to being 500 terms. Upon observing the results, we noticed that Madison appeared slightly more frequently than Hamilton. Looking at the Naive Bayes score for each document, it was apparent that Hamilton and Madison had close scores near each other. Some scores were 0.02 - 0.002 point difference. We do see that Madison did occur more times than Hamilton, yet the data should have shown Madison a lot more. With comparison to the Rocchio data set though, the prediction of Madison on the Rocchio classifier, mostly agrees with the Bayesian procedure.

Document Number	Rocchio	Bayesian
49	Jay	Madison
50	Madison	Hamilton
51	Madison	Madison
52	Madison	Madison
53	Madison	Hamilton
54	Madison	Hamilton
55	Jay	Hamilton
56	Madison	Hamilton
57	Jay	Madison
62	Madison	Madison
63	Madison	Madison

VI. Conclusion

Previous statistical analysis has attributed all disputed papers above to Madison. It is interesting to see how our classifiers faired in trying to correctly classify these papers. In regards to the Bayesian Classification method, we feel that our implementation can be improved. Our output did not correctly classify each contested document as Madison. We believe that some of the calculations are wrong, which in turn produced the wrong output. Rocchio classification fared better, with 8 out of the 11 documents being correctly classified for Madison. Given that Rocchio classification is “inaccurate if classes are not approximately spheres with similar radii”, and that some of these papers were probably collaborative efforts, we feel a 27% error is within reason. (Christopher D. Manning 2009) Further, natural language experts composed a rhetorical analysis of the disputed documents, and the corpus as a whole, and claim that the documents do contain a polyphony, or multiple voices and identities. (Jasinski 1997)

Bibliography

Adair, Douglass. 1944. "Authorship of Disputed Federalist Papers." *The William and Mary Quarterly* 97-122.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2009. *An Introduction to Information Retrieval.* Cambridge, England: Cambridge University Press.

Jasinski, James. 1997. "Heteroglossia, Polyphony, and "The Federalist Papers"." *Rhetoric Society Quarterly, Vol. 27, No 1* 23-46.