# *The Federalist Papers* Revisited: A Collaborative Attribution Scheme

**Jacques Savoy**
Computer Science Dept., University of Neuchatel
2000 Neuchatel (Switzerland)
Jacques.Savoy@unine.ch

## ABSTRACT

This paper presents and evaluates a collaborative attribution strategy based on six authorship attribution schemes representing the two main paradigms used in authorship studies. Based on very frequent words as features, the classical paradigm (or similarity-based methods) proposes to compute an intertextual distance between the disputed text and the different author profiles (concatenation of their writings). As a second paradigm, we can apply different machine learning schemes such as the naïve Bayes, and the support vector machines (SVM). As an evaluation corpus, we have used *The Federalist Papers*, a well-known collection in authorship attribution. During our evaluation, we have tried to follow the recommendations and the best practices known to assess the various attribution schemes. The evaluation shows that, in the two paradigms, we can find effective attribution schemes. But when combining these individual results using a vote aggregation method, the final collaborative decision is always correct and robust. Moreover, to indicate the degree of belief attached to the combined attribution, we can consider the percentage of votes obtained by each possible assignment. When analyzing the output given by the individual attribution schemes, we also found that the provided information is difficult to interpret, at least, for the end-user.

### Keywords

Authorship Attribution; Stylometry; Text Categorisation; Federalist Papers.

## INTRODUCTION

In this paper, we address the authorship attribution (AA) problem (Juola, 2006) whereby the author of a given text must be determined based on text samples written by known authors. Knowing that the real author is one of the candidates, this specific challenge is defined as the *closed-class* AA problem. In such applications, the disputed text might correspond to various items such as a part of a romance, an anonymous letter, a Web page, an e-mail, a blog post, etc.

In classical AA studies, we usually focus on frequent words to represent each text. The underlying idea is to assume that the words used often and unconsciously vary from one author to another. Thus they are able to reveal the individual "fingerprint" of the real writer. An intertextual distance measure can then be defined based on those selected terms. To determine the most probable writer we select the author profile depicting the smallest distance with the disputed text.

As an alternative way, the AA problem can be viewed as a categorization problem in which each author corresponds to one category. In this case, based on a training set, a machine learning scheme can learn the distinctive fingerprints of the various authors. After representing the disputed text based on the selected features, the classifier can determine the most probable or most unlikely writer.

Recently, new attribution schemes have been proposed and evaluated using different test collections. Moreover, the evaluation results were reported after a fine-tuning of the underlying parameters. The question that then arises is the following: which attribution scheme is the most effective when using the same evaluation corpus and the default parameter setting? To answer this question, we have based our experiment on *The Federalist Papers*, a set of articles written to persuade the people of New York to adopt the US Constitution. Based on the results given by six single attribution schemes, we demonstrate that a vote aggregation method results in a more robust, reliable and correct decision.

The rest of this paper is divided as follows. The next section presents related work and the attribution schemes used in our experiments. The third section outlines the main characteristics of *The Federalist Papers* while our evaluation and collaborative scheme are described in the fourth section. Finally the last section presents and analyses the information provided by some classifiers to justify their choices.

## RELATED WORK

To solve the AA problem, classical attribution schemes are based on the idea of measuring an intertextual distance based on the vocabulary used. In this vein, Mosteller & Wallace (1964) proposed to manually select the

*ASIST 2013,* November 1-6, 2013, Montreal, Quebec, Canada.

most frequent and useful terms composed mainly by function words (determiners, prepositions, conjunctions, pronouns and some adverbs and verbal forms). In their final study, this list contains a reduced set of 35 terms.

Following this vein, Burrows (2002) suggested automatically selecting word types able to discriminate between authors by considering their occurrence frequencies. To compute a distance between two texts, Burrows proposed evaluating standardized term frequencies. To achieve this, a Z score value is computed for each term $t_i$ by calculating its relative term frequency $rtf_{ij}$ in a document $D_j$, as well as the mean ($mean_i$), and standard deviation ($sd_i$) of term $t_i$ over all texts belonging to the corpus. For each term, we compute a Z score ($t_{ij}$) = ($rtf_{ij}$ - $mean_i$) / $sd_i$.

Based on these quantities, we can then compute the distance between a query text Q and each author profile $A_j$ (concatenation of all texts written by the same writer). Given a set of terms $t_i$, for $i$ = 1, 2, …, $m$, the Delta value (denoted Δ) is computed according to Equation 1. When comparing a disputed text with different author profiles, the lowest Δ measure indicates the most probable author.

$$\Delta(Q,A_j) = {}^{1}\!/_{m} \cdot \sum_{i=1}^{m} \left| Z\,score(t_{iq}) - Z\,score(t_{ij}) \right| \qquad (1)$$

As another attribution scheme, Grieve (2007) considered selecting all words in a $k$-limit profile, where $k$ indicates that each selected term must occur, at least, in $k$ documents written by every author (e.g., a value $k = 5$ imposes the presence of the target word in at least five texts written by every possible author). This scheme imposes that all selected terms must be used by all authors, and the best $k$ values seem to lie between 2 and 5.

After this term selection procedure, Grieve (2007) uses the chi-square statistic defined by Equation 2 to compare a given query text Q with an author profile $A_j$. In this formulation, $rtf_{iq}$ represents the relative frequency of the $i$th term in the query text Q, $rtf_{ij}$ the relative frequency in the $j$th author profile $A_j$, and $m$ the number of selected terms $t_i$ in a $k$-limit.

$$\chi^2(Q,A_j) = \sum_{i=1}^{m} {\left( rtf_{iq} - rtf_{ij} \right)^2} \Big/ {rtf_{ij}} \qquad (2)$$

The lowest chi-square value is used to determine the most probable author, or these values can be used to rank the different possible authors.

Zhao & Zobel (2007) propose to define *a priori* the most useful word types. Their suggested list contains 363 English word types, composed mainly of function words but with some lexical terms (but independent of the topics of the underlying texts). This approach owns the advantage to be independent of the underlying corpora and can be applied with various classification strategies. To compute the distance between two text representations, Zhao & Zobel (2007) propose using the Kullback-Leibler divergence (KLD).

$$KLD(Q \parallel A_j) = \sum_{i=1}^{m} \mathrm{Prob}_{\mathrm{q}}\left[t_i\right] \cdot \log_2 \left[ \frac{\mathrm{Prob}_{\mathrm{q}}\left[t_i\right]}{\mathrm{Pr\,ob}_{\mathrm{Aj}}\left[t_i\right]} \right] \qquad (3)$$

To estimate the probability of having the corresponding term $t_i$ in the query text ($\mathrm{Prob}_q[t_i]$) or in the author profile ($\mathrm{Prob}_{Aj}[t_i]$), we may consider the term's absolute occurrence frequency (denoted $tf_i$) and the size of the corresponding text ($n$) (e.g., Prob[$t_i$] = $tf_i/n$). Usually it is better to smooth these estimates, and we have applied the Lidstone's technique (Manning & Schütze, 1999) to obtain Prob[$t_i$] = ($tf_i$+λ) / ($n$+λ·|V|), with |V| indicating the vocabulary size, and λ = 0.1 (producing usually slightly better performance over other choices).

As a fourth authorship attribution approach, we suggest representing each text based on selected terms corresponding to its specific vocabulary (Savoy, 2012). To measure a word's specificity in a part $P_0$, we consider its absolute occurrence frequency in $P_0$ (denoted $tf_{i0}$), and its occurrence frequency ($tf_{i1}$) in the rest of the corpus (denoted $P_1$). For the whole corpus ($P_0 \cup P_1$) the absolute occurrence frequency of the term $t_i$ is $tf_{i0}$+$tf_{i1}$. The total number of tokens in $P_0$ is denoted $n_0$, while the size of the whole corpus is given by $n$.

The distribution of each term $t_i$ in $P_0$ is assumed to follow a binomial with parameters $n_0$ and Prob[$t_i$] (the probability of selecting the term $t_i$ from the corpus, estimated as Prob[$t_i$] = ($tf_{i0}$+$tf_{i1}$) / $n$). A good practice, however, is to smooth such estimates (Lidstone's smoothing with λ = 0.1) (Manning & Schütze, 1999).

Repeating this drawing $n_0$ times we are able to estimate the expected number of occurrences of term $t_i$ in $P_0$ by $n_0$ · Prob[$t_i$]. Then we can compare it to the observed number (namely $tf_{i0}$), and any large difference between these two values indicates a deviation from the expected behavior. To have a more precise definition of *large*, we account for the variance (in a binomial process, it is defined as $n_0$ · Prob[$t_i$] · (1-Prob[$t_i$])). Equation 4 defines the standardized score for $t_i$ in $P_0$.

$$\text{Z score}(t_{i0}) = \frac{tf_{i0} - n_0 \cdot \text{Prob}[\text{t}_\text{i}]}{\sqrt{n_0 \cdot \text{Prob}[\text{t}_\text{i}] \cdot (1 - \text{Prob}[\text{t}_\text{i}])}} \qquad (4)$$

Such a Z score is assigned to each term $t_i$. From these values we define the distance between a query text Q and an author profile $\text{A}_\text{j}$ as defined by Equation 5. In this formula, $t_{iq}$ indicates the *i*th term in the query text, $t_{ij}$ indicates the *i*th term in the *j*th author profile $\text{A}_\text{j}$, and *m* the number of selected terms.

$$\text{Dist}(Q, A_j) = \frac{1}{m} \cdot \sum_{i=1}^{m} \left(\text{Z score}(t_{iq}) - \text{Z score}(t_{ij})\right)^2 \qquad (5)$$

Besides these four classical approaches, we can follow a second main paradigm based on machine learning techniques applied in text categorization problems, or in authorship attribution (Stamatatos, 2009). In this view, we see each author as one possible category. As a learning scheme, we selected the naïve Bayes model (Manning *et al.*, 2008) to determine the probable writer between the set of possible authors, denoted by $\text{A}_\text{j}$ for j = 1, 2 , … *r*. To define the most probable author of a query text Q, the naïve Bayes model selects the one maximizing Equation 6, in which $t_i$ represents the *i*th term included in the query text Q, and $n_q$ indicates the size of the query text.

$$\text{Arg}\max_{\text{Aj}} \text{Prob}[A_j \mid Q] \propto \text{Prob}[A_j] \cdot \prod_{i=1}^{n_q} \text{Prob}[t_i \mid A_j] \qquad (6)$$

To estimate the prior probabilities of each author ($\text{Prob}[\text{A}_\text{j}]$), we can assigned the same chance to each possible author (uniform distribution or uninformed prior). As a variant, we can simply take into account the proportion of articles written by each author. To determine the term probabilities $\text{Prob}[t_i \mid \text{A}_\text{j}]$, we regroup all texts belonging to the same author to define the author profile. For each term $t_i$, we then compute the ratio between its absolute occurrence frequency in the author profile $\text{A}_\text{j}$ ($tf_{ij}$) and the size of this sample ($n_j$). As with the previous methods, we apply the Lidstone's smoothing method and defined $\text{Prob}[t_i \mid \text{A}_\text{j}] = (tf_{ij}+\lambda) / (n_j+\lambda \cdot |\text{V}|)$ (with $\lambda = 0.1$).

As a second approach, based on the machine learning paradigm, we used the SVM model (Joachims, 2001) which usually performed well on various text categorization tasks (Joachims, 2002). In this model, each text is represented by a term vector. To reflect the importance of each term in this representation, we define a term weight. Derived from the vector-space model, a well-known technique is to weight each term through applying the *tf idf* formula (Joachims, 2002) (Manning *et al.*, 2008), in which the component *tf* represents the number of occurrences within the text. The *idf* (= log(*df* / *n*)) corresponds to the logarithm of the inverse document frequency (denoted *df*), and thus indicates the number of texts in which the corresponding term occurs, while *n* indicates the total number of texts in the corpus. Such weighting schemes have been found effective in various text classification tasks (Joachims, 2002).

## THE FEDERALIST PAPERS

In authorship attribution, *The Federalist Papers* (Rossiter, 2003) represents a corpus composed of 85 articles from which twelve are disputed between two possible authors (Mosteller & Wallace, 1964). This corpus was written to persuade "the People of the State of New York" to ratify the Constitution (Maier, 2010). These papers were published (and republished) between October 1787 and May 1788 in newspapers under the pseudonym of *Publius*. Under this name, General Alexander Hamilton (1755-1804), James Madison (1751-1836) and John Jay (1745-1829) have jointed their efforts to present the merits of the new Constitution and to answer critics formulated by the Anti-federalist (Ketcham, 2003).

| # | Benson | Madison | Current |
|---|---|---|---|
| 1 | Hamilton | Hamilton | Hamilton |
| 2-5 | Jay | Jay | Jay |
| 6-9 | Hamilton | Hamilton | Hamilton |
| 10 | Madison | Madison | Madison |
| 11-13 | Hamilton | Hamilton | Hamilton |
| 14 | Madison | Madison | Madison |
| 15-17 | Hamilton | Hamilton | Hamilton |
| 18-20 | Mad. & Ham. | Madison | Mad. & Ham. |
| 21-36 | Hamilton | Hamilton | Hamilton |
| 37-48 | Madison | Madison | Madison |
| 49-53 | Hamilton | Madison | Madison |
| 54 | Hamilton (Jay) | Madison | Madison |
| 45-58 | Hamilton | Madison | Madison |
| 59-61 | Hamilton | Hamilton | Hamilton |
| 62-63 | Hamilton | Madison | Madison |
| 64 | Jay (Hamilton) | Jay | Jay |
| 65-85 | Hamilton | Hamilton | Hamilton |

**Table 1. Authorship of the 85 *Federal Papers* according to Benson's list, Madison, and the current attribution**

If, at the time of publication, the authorship of each paper was kept secret, contemporaries have guessed the joint work of Hamilton, Madison and Jay, without being able to explicitly attribute each given paper to its legitimate author. In 1804, two days before his fatal duel, Hamilton gave the first assignment (Benson's list, see Table 1). In this list, there is a large consensus agreeing that a substitution occurs between the author's name of Paper #54 and #64. We will adopt this position and admit that

Hamilton wrote Paper #54 (instead of Jay as specified in the Benson's list) while Jay is the author of Paper #64 (instead of Hamilton). After his presidency in 1818, Madison gives his assignment, revealing 15 differences between the two lists. The last column of Table 1 indicates the current admitted attribution of each *Federalist* paper. This position reflects a large consensus but some authors don't share this attribution as, for example, Rudman (2012) who suggests that the disputed papers are jointly written by Hamilton & Madison.

In this table, we can see that 70 articles are undisputed (5 by Jay, 14 by Madison, and 51 by Hamilton). From this set we will ignore the five articles written by Jay (Paper #2 to #5, and Paper #64) and the three papers written jointly by Hamilton and Madison (Paper #18 to #20). We can mention that the limited contribution of Jay could be explained by his illness during the winter 1787-88.

The undisputed 65 articles will form the training set used to define the stylistic characteristics of the two possible authors. In the test set, we count twelve disputed papers that could have been written by either Hamilton or Madison (Paper #49 to #58 and #62 to #63).

To illustrate the difficulty of correctly attributing the disputed articles, we analyzed the occurrence frequency of the various Part-Of-Speech (POS) in the training set. When comparing the percentage of determiners used in Hamilton's articles vs. Madison's writings, the difference is rather small (17.78% by Hamilton vs. 17.88% by Madison, a difference of 0.1%). On the other hand, when comparing Hamilton's with Jay's papers, the difference is larger and rises to 5.8% (17.78% vs. 11.95% for Jay). For nouns, we detected a similar pattern. The difference in percentage between Hamilton (22.8%) and Madison (22.2%) is small (0.6%), while with Jay's articles (20.6%) the difference is larger (2.2%). The largest difference between Hamilton and Madison appears with the word *to* used to indicate the infinitive (Hamilton: 4.4%, Madison: 3.5%, difference: 0.9%; Jay: 3.7%). Clearly the Hamilton's and Madison's styles are closely related, while Jay's style is more distant, according to the percentages of each POS tag.

This corpus is also interesting for historical reasons. In fact, this set of commentary papers on the principles of government is still an important source of interpretation for the US Constitution (Rossiter, 2003), (Meyerson, 2008). From an AA perspective, this sample is also useful because it is formed of texts having the same overall topics, genre, intent, and that are extracted from the same time period. Previous studies in AA have shown that the style differs from one person to another but it is also influenced by the period, the topics, the genre, and the text intention (Juola, 2003), (Labbé, 2007), (Hughes *et al.*, 2012).

To generate this corpus, we have downloaded *The Federalist* from the Gutenberg project. All the text was then transposed to lowercase and tokenized to determine words (sequence of letters or digits) and punctuation symbols. This pre-processing is relatively simple, ignoring, for example, the Part-Of-Speech information.

**EVALUATION**

Table 2 depicts the evaluation results of the six selected attribution schemes using the twelve disputed articles from *The Federalist* (for which the "correct" assignment is to attribute them to Madison). The first row indicates the attribution model and the number of selected terms.

As a first method, we used the Delta rule (Burrows, 2002) based on the 50 most frequent terms. This approach produces two "errors" (Paper #55, and #56). With this method, the number of suggested terms may vary from 40 to 200 (Burrows, 2002). In our case, we applied the default setting, namely the 50 most frequent words.

As the second AA method, we have evaluated the chi-square metric (Grieve, 2007) based on 2-limit selection (each selected term must appears in, at least, two articles written by each possible author). This parameter setting produces the best performance in (Grieve, 2007). In our experiment, the system extracts 1,177 terms and this attribution scheme produces one "error" (Paper #56).

As a third attribution strategy, we used the KLD scheme proposed by Zhao & Zobel (2007) based on a predefined list of 344 English words. The single parameter used in this approach is the λ value (Lidstone's smoothing, fixed at 0.1). This strategy renders a perfect answer.

With the Z-score method (specific vocabulary), each selected term must be used by all possible authors and have a high occurrence frequency (higher than 300 in our experiment) (Savoy, 2012). By applying this selection, we extracted 75 terms and a high accuracy rate was obtained.

| Papers | Delta | Chi-square | KLD | Z-score | Naïve Bayes | SVM |
|---|---|---|---|---|---|---|
| | 50 terms | 1,177 terms | 344 terms | 75 terms | 344 terms | 344 terms |
| #49 | Madison | Madison | Madison | Madison | Madison | Madison |
| #50 | Madison | Madison | Madison | Madison | Madison | Madison |
| #51 | Madison | Madison | Madison | Madison | Madison | Madison |
| #52 | Madison | Madison | Madison | Madison | Madison | Madison |
| #53 | Madison | Madison | Madison | Madison | Madison | Madison |
| #54 | Madison | Madison | Madison | Madison | Madison | Madison |
| #55 | *Hamilton* | Madison | Madison | Madison | Madison | Madison |
| #56 | *Hamilton* | *Hamilton* | Madison | Madison | Madison | Madison |
| #57 | Madison | Madison | Madison | Madison | Madison | *Hamilton* |
| #58 | Madison | Madison | Madison | Madison | Madison | *Hamilton* |
| #62 | Madison | Madison | Madison | Madison | Madison | Madison |
| #63 | Madison | Madison | Madison | Madison | Madison | Madison |

**Table 2. Authorship attribution for the twelve disputed *Federalist Papers***

Before applying the naïve Bayes model, we need to define a set of discriminative terms. To simplify this feature selection procedure, we will use the words appearing in Zhao's list (344 terms). With this set of frequently used words, we obtain a perfect accuracy rate of 100% when using, as prior, an uniform distribution (both authors have the same probability to be the real author of each disputed text). However, when considering that Hamilton wrote 51 papers and Madison only 14 articles, this information can be used to define a new prior distribution (Hamilton: 51 / (51+14) = 0.78; Madison: 14 / (51+14) = 0.22). Using this prior distribution favoring clearly the most frequent author, the naïve Bayes model assigns all disputed papers to Hamilton.

When using the SVM method, we need to select a set of discriminative features. As for the naïve Bayes approach, we have used Zhao's list (344 terms). To weight these words in each text representation, we have used the well-known *tf idf* text representation (Manning *et al.*, 2008). Based on this representation, we used the available SVM package (`kernlab`) (Karatzoglou *et al.*, 2006) for the R language (Crawley, 2007). With this attribution scheme, we observe two "errors" (Paper #57, and #58). In this evaluation, the parameter *c* (cost of a misclassification) was set to 1 (the default value), and we applied a linear kernel (default option). Of course other feature selection procedures and weighting schemes can be applied within the SVM model. As an extreme example, Fung (2003) shows that we can use only three terms (namely *to*, *upon*, and *would*) to assign all disputed papers to Madison.

Overall, the most effective attribution schemes are the KLD (Zhao & Zobel, 2007), the Z-score (Savoy, 2012), and the naïve Bayes model (Manning *et al.*, 2008) but, in the latter case, with an uninformed prior distribution. The chi-square measure produces one "error", while the SVM, or the Delta approach generate two "errors." When making slight variations to the parameter values, we achieve similar results.

The evaluation results reported in Table 2 are valid for *The Federalist* corpus. When having a new single disputed text, which AA model will perform the best? Are we sure that the KLD, the Z-score or the naïve Bayes will always return the correct answer?

Of course, the answer is 'no'. But when having multiple experts (or multiple attribution models), we can take into account all of the opinions by combining them. In fact, each attribution scheme uses a different selection procedure generating different sets of selected terms. Those features are then weighted and combined according to different classification algorithms. Thus each attribution scheme tends to account for different stylistic elements that may result in different attribution decisions.

As a simple aggregation procedure, we suggest to adopt a voting method in which each expert (or proposed assignment) owns the same importance. Based on the result depicted in Table 2, we can see that this approach will achieve a perfect answer. In fact, the most difficult case is Paper #56 with two votes for Hamilton and four for Madison. This is not a surprise because it is known that this paper is rather problematic to assign with a high degree of certainty.

We must mention that such a collaborative decision strategy also has the advantage of being more robust than considering a single attribution scheme. In general, a difficult AA case might be incorrectly classified by an attribution scheme taking its decision on a reduced set of stylistic features. But when considering several attribution models the final decision is more robust when considering possible noisy term selection, different weighting and classification procedures.

Finally, such an aggregation method offers an indication about its degree of certainty (or belief) about the proposed assignment. This degree is correlated to the percentage of votes obtained by the most frequent answer. In Table 2, we can observe that all the attribution schemes propose Madison as the real author of Paper #49, #50, #51, #52, #53, #54, #62, and #63. On the other hand, the degree of certainty is lower for Paper #56. A more detailed analysis of this last attribution is needed, and this is the aim of the next section.

## DEEPER ANALYSIS

The evaluations reported in the previous section can only be achieved, however, when we know the correct decisions for each disputed text. In practice, the correct answer will often stay unknown, and estimating the accuracy rate under new conditions is always problematic (Hand, 2006).

When applying some classifiers, the produced output could be limited to the most probable author's name. This is the default output with the SVM package that we used. In such circumstances, it is impossible to have an idea of the closeness of other possible authors. With other attribution schemes, we may obtain a ranked list of possible authors. Table 3 depicts an extended output provided by the various schemes for Paper #56.

| Method | Ranked list |
|---|---|
| Delta | 1. Hamilton: 1.006<br>2. Madison: 1.022 |
| C | 1. Hamilton: 2.209<br>2. Madison: 2.630 |
| KLD | 1. Madison: 0.203<br>2. Hamilton: 0.248 |
| Z-score | 1. Madison: 9.833<br>1. Hamilton: 10.767 |
| Naïve Bayes | 1. Madison: -4.755<br>2. Hamilton: -4.786 |
| SVM (*tf idf*) | Madison: 0.241 |
| SVM (*tf idf*) | 1. Madison: 91.1%<br>2. Hamilton: 8.9% |

**Table 3. Extended output of different attribution schemes for Paper #56**

As shown, each classifier usually provides a ranked list of the two author's names with a numerical indication reflecting the fitness of the corresponding author's profile to Paper #56. This value corresponds to the estimated textual distance for the Delta rule (see Equation 1), the chi-square statistic (Equation 2), the measure of the Kullback-Leibler divergence (Equation 3), and the intertextual distance for the Z-score approach (Equation 5). For the naïve Bayes model, the reported value is proportional to the logarithm of the probability of being the right author (the logarithm of Equation 6). Finally, the SVM algorithm may return the distance of the disputed text to the hyperplane defining the border between the two classes (e.g., 0.241 in this case). As an alternative output, the estimated probability for both possible authors can be obtained (as depicted in the last row of Table 3). These numerical values (fitness) are used as a key to sort the two possible authors. As we can see, their magnitude and differences are usually difficult to interpret, at least for the end-user. For example, based on the Delta rule, how can we interpret the difference (0.016) between Hamilton's (1.006) and Madison's attribution (1.022)? Should we consider this difference as large or small? Moreover, the comparison between the fitness computed by different disputed texts is difficult to interpret, in part because the disputed articles do not have the same size.

A second major concern is the capability of the attribution scheme to provide a reason explaining the proposed assignment. A numerical value or a difference between two distances does not convey a pertinent justification.

| | **Hamilton** | | **Madison** |
|---|---|---|---|
| 1. the | 0.1326 | 1. the | 0.1421 |
| 2. , | 0.0967 | 2. , | 0.1028 |
| 3. of | 0.0921 | 3. of | 0.0843 |
| 4. to | 0.0580 | 4. to | 0.0456 |
| 5. . | 0.0386 | 5. and | 0.0388 |
| 6. in | 0.0358 | 6. . | 0.0294 |
| 7. and | 0.0345 | 7. in | 0.0280 |

**Table 4. The seven terms having the highest relative frequencies in both author's profiles**

In order to explain a proposed attribution, we can inspect the relative frequencies (or the occurrence probabilities) associated with each term. After sorting terms according to their decreasing occurrence frequencies, the most frequent ones tend to appear in a similar order for both author's profiles. For example, Table 4 depicted the seven terms depicting the highest relative frequencies based on the *Federalist* corpus. As we can see, the ranking is similar, only the relative frequencies (or probability estimates based only the 100 most frequent word types) differ from one author to another. In this example, the four most frequent terms appear in the same order in both profiles.

From such similar pattern, a stylistic interpretation and an overall understandable picture are thus not easy to derive.

A better interpretation can be based on the specific vocabulary (Savoy, 2012). With this model, we establish the word types each possible author tends to over-use or under-use. Based on the 75 most frequent terms in the training set, Table 5 shows the top six most over-used terms and the three most under-used terms for the two possible authors.

| | Hamilton | Madison |
|---|---|---|
| Over-used terms | upon<br>would<br>to<br>there<br>courts<br>kind | powers<br>confederation<br>department<br>on<br>congress<br>and |
| Under-used terms | on<br>representatives<br>by<br>department | upon<br>there<br>would<br>to |

**Table 5. Terms more over-used or under-used by the two possible authors of the *Federalist***

Based on the information shown in Table 5, we can see that Hamilton tends to over-use the prepositions *upon*, *to*, the verbal form *would,* or the nouns *courts* and *kind*. He also clearly under-uses the prepositions *on* and *by,* or the nouns *representatives* or *department*. This last term is over-used by Madison who also prefers using the conjunctions *and*, the preposition *on*, and the lexical forms *powers*, *confederation*, and *congress*.

| | Paper #56 | Madison | | Hamilton | |
|---|---|---|---|---|---|
| | rtf | rtf | Z-score | rtf | Z-score |
| the | 0.1346 | 0.1544 | 3.73 | 0.1430 | -2.42 |
| , | 0.1137 | 0.1117 | 2.69 | 0.1043 | -1.89 |
| of | 0.1117 | 0.0916 | -2.66 | 0.0993 | 2.45 |
| and | 0.0528 | 0.0462 | 5.58 | 0.0372 | -2.92 |
| a | 0.0479 | 0.0305 | -2.63 | 0.0344 | 1.38 |
| to | 0.0399 | 0.0495 | -5.54 | 0.0625 | 5.52 |
| be | 0.0369 | 0.0298 | -1.93 | 0.0315 | 1.10 |
| in | 0.0299 | 0.0320 | -3.86 | 0.0386 | 3.30 |

**Table 6. Relative term frequencies (rtf) in Paper #56 and in the two author profiles, with their Z-score values**

In Table 6, we have reported the relative frequencies of some frequently occurring terms in Paper #56. In addition, we have depicted the relative frequencies in Madison and Hamilton's profiles. Finally, Table 6 indicates the Z-score values in both author's profiles (based on the 75 terms selected by the Z-score attribution scheme). Based only on these terms and their relative frequencies (*rtf*), an assignment seems rather difficult to derive and problematic to justify.

When considering the Z-score values computed according to both author's profiles, a simple assignment seems possible when considering one or a few terms. The first two words are over-used in Madison's writings (Z-score values positive in Madison's profile). The same pattern occurs with the conjunction *and*. These three terms indicate a possible attribution to Madison. On the other hand, the prepositions *of* and *to* are over-used by Hamilton (Z-score values positive in Hamilton's profile). Moreover, the occurrences of the word types *a*, *be*, or *in* tend to favor a possible attribution to Hamilton. Thus, each possible attribution has evidence in its favor.

In conclusion, this example demonstrates the real difficulty of assigning the right author to a given text. When considering the output of several attribution models, a combined decision will smooth the weight attached to stylistic features specific to each single attribution scheme.

## CONCLUSION

In this paper we analyzed and evaluated the authorship attribution problem with four similarity-based attribution schemes. In this paradigm, we selected very frequent word types as style markers. It is assumed that such terms are not fully under the control of the author and can thus be appropriate features for discriminating among different writers.

As a second paradigm, we can also view the AA problem as a specific task in automatic text categorization. Using a predefined list of frequent terms as features, we evaluated two classifiers, namely the naïve Bayes, and the SVM approach.

Using *The Federalist* as an evaluation corpus, our experiments show that some approaches in both paradigms can provide good overall performance. In particular, the KLD and Z-score scheme in the classical family, or the naïve Bayes in the machine learning domain provide the best overall results.

We must recognize however that the output produced by a classification algorithm is usually difficult to interpret. In practical applications, we are often faced with a single disputed text. In such cases, when an attribution scheme returns an intertextual distance, a correct and useful interpretation of this measure is rather difficult to derive. On the other hand, when a probability estimate is provided, its underlying variability is unknown, rendering the interpretation of this probability quite problematic as well.

As a collaborative attribution scheme, we suggest to adopt a simple voting method in which each single attribution scheme has the same importance. In this case, we can take account of different stylistic features detected by different attribution models. Even working with the same set of features, distinct attribution models will weight them differently and compute a classification decision based on different algorithms. The final decision obtained by the majority of the attribution schemes tends to be more robust. Moreover, the percentage of votes for the winner can provide an indication about the degree of belief or certainty about the proposed decision. When the resulting percentage is high for a given attribution, we have a set of corroborating evidence in favor of one author (without having an definitive and absolute certainty). In this perspective, this study confirms Madison as the real author of *Federalist Paper* #49, #50, #51, #52, #53, #54, #62, and #63.

**REFERENCES**

Burrows, J.F. (2002). Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary & Linguistic Computing*, 17(3), 267-287.

Crawley, M.J. (2007). *The R Book*. Chichester: John Wiley & Sons.

Fung, G. (2003). The Disputed *Federalist Papers*: SVM Feature Selection via Concave Minimization. *Proceedings TAPIA-2003* (pp. 42-46). New York: The ACM Press.

Grieve, J. (2007). Quantitative Authorship Attribution: An Evaluation of Techniques. *Literary & Linguistic Computing*, 22(3), 251-270.

Hand, D.J. (2006). Classifier Technology and the Illusion of Progress. *Statistical Science*, 21(1), 1-14.

Hughes, J.M., Foti, N.J., Krakauer, D.C., and Rockmore, D.N. (2012). Quantitative Patterns of Stylistic Influence in the Evolution of Literature. *Proceedings of the PNAS,* 109(20), 7682-7686.

Joachims, T. (2001). A Statistical Learning Model of Text Categorization for Support Vector Machine. In *Proceedings of ACM SIGIR '2001* (pp. 128-136). New York: The ACM Press.

Joachims, T. (2002). *Learning to Classify Text Using Support Vector Machines. Methods, Theory and Algorithms*. London: Kluwer.

Juola, P. (2003). The Time Course of Language Change. *Computers and the Humanities*, 37(1), 77-96.

Juola, P. (2006). Authorship Attribution. *Foundations and Trends in Information Retrieval*, 1(3).

Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support Vector Machines in R. *Journal of Statistics Software*, 15(9), 1-28.

Ketcham, R. (Ed.). (2003). *The Anti-Federalist Papers and the Constitutional Convention Debates*. New York (NY): Signet Classics.

Labbé, D. (2007). Experiments on Authorship Attribution by Intertextual Distance in English. *Journal of Quantitative Linguistics*, 14(1), 33-80.

Maier, P. (2010). *Ratification. The People Debate the Constitution, 1787-1788*. New York (NY): Simon & Schuster.

Manning, C.D., & Schütze H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge (MA): The MIT Press.

Manning, C.D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Meyerson, M.I. (2008). *Liberty's Blueprint. How Madison and Hamilton Wrote the Federalist Papers, Defined the Constitution, and Made Democracy Safe for the World*. Philadelphia (PA): Basic Books.

Mosteller, F., & Wallace, D.L. (1964). *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. Reading (MA): Addison-Wesley.

Rossiter, C. (Ed.). (2003). *The Federalist Papers*. New York (NY): Signet Classics.

Rudman, J. (2012). The Twelve Disputed '*Federalist*' Papers: A Case for Collaboration. *Proceedings Digital Humanities 2012*, 353-356.

Savoy, J. (2012). Authorship Attribution Based on Specific Vocabulary. *ACM – Transactions on Information Systems*, 30(2).

Stamatatos, E. (2009). A Survey of Modern Authorship Attribution Methods. *Journal American Society for Information Science and Technology*, 60(3), 433-214.

Zhao, Y., & Zobel, J. (2007). Entropy-Based Authorship Search in Large Document Collection. *Proceedings ECIR-2007* (pp. 381-392). Berlin: Springer-Verlag.