CEUR-WS.org/Vol-3313/paper5.pdf

Demystifying the scribes behind the Voynich Manuscript using Computational Linguistic Techniques

Kevin Farrugia, Colin Layfield and Lonneke van der Plas

University of Malta, Msida, MSD 2080, Malta

Abstract

Earlier work studies the palaeography of the Voynich manuscript and proposes five scribes. In this work, we use computational linguistic models that are based on the character sequences found on the pages, inspired by work on automatic authorship attribution, as opposed to the earlier work focusing on the nature of the writing script. Our rationale is that if these two independent methods should lead to similar results, we have strong evidence for the earlier identification of the scribes in the Voynich manuscript. We carry out a machine learning experiment where the manuscript is divided into an equal number of pages for three of the five scribes. Four machine learning classifiers create models that have been trained on the classification produced in earlier work to predict a scribe for each page using character n-grams as features. The results of each classifier are analysed comparatively with one another and compared with the classification provided based on palaeographic work. The results show when the classifiers produce the same classification as earlier palaeographic work on a held-out dataset.

Keywords

Voynich manuscript, scribes, machine learning, stylometric features, cross-validation

1. Introduction

The Voynich manuscript, kept in Yale University's Beinecke Rare Book and Manuscript Library, is a medieval manuscript which has baffled medievalists, cryptographers, and linguists for almost a century as, to this day, the cryptic characters that make up words in this codex are from an unknown script. No other examples of works using the same language as the text in the manuscript are known. The intent of creating a piece of writing which cannot be understood by others of contemporary and future generations seems to contrast with why people primarily write – to exchange information. This is the reason why there isn't universal agreement on many of the book’s elements [2].

Earlier work by Davis [1] identified five scribes or hands in the Voynich manuscript based on palaeographic investigations. The goal of this study is not to create a new classification for the potential scribes who created the document using machine learning, but rather to analyse Davis's suggested classification [1] and evaluate it against itself using computational linguistic techniques. It aims to highlight those pages in her categorisation that were possibly misclassified. Since Davis's findings are the most recent and comprehensive ones that can be found regarding the number of scribes and their distribution of work, they are used as the basis for this, thus creating a baseline for comparison. The research questions this paper aims to answer are as follows:

1. Does the classification of pages into scribes, as proposed in palaeographic works, hold out when cross-validated by means of automatic authorship attribution?

International Conference on the Voynich Manuscript 2022, November 30--December 1, 2022, University of Malta.
EMAIL: kevin.j.farrugia.15@um.edu.mt (A. 1); colin.layfield@um.edu.mt (A. 2); lonneke.vanderplas@idiap.ch (A. 3)
ORCID: 0000-0002-1868-4258 (A. 2)
© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

In the process of answering the above, we will investigate the following:

2. What components of the undeciphered text of the Voynich manuscript are adequate to use as features that determine one scribe from the other?
3. Which pages of the manuscript prove to be more difficult than others to categorise into scribes as proposed by the findings in [1]?

2. Previous Work

2.1. About the Voynich Manuscript

This paper is particularly interested in how the manuscript's content is divided up, as it looks at page classification into scribes. The manuscript itself is roughly 23 by 16 cm and is 5 cm thick [3]. It is made up of bi-folios, and four of them combined form a quire. The bifolios, as their name implies, are composed of two folios that each have a page on both their recto and verso sides (shortened in notation as r, v). These folios add up to 102 in total, however the manuscript originally had 116 folios. There are 18 quires, the majority of which are made up of 4 bifolios, which themselves are made up of 8 folios, with a total of typically 16 pages each quire [4]. This arrangement has several exceptions, such as quire 8, which appears to have had five bifolios before some of them were lost to history [1]. According to linguistic analysis, the manuscript contains around 35,000 words or 170,000 characters [5]. No sign of punctuation is present [6].

A variety of themes have been discovered once the drawings were analysed, splitting the Voynich manuscript into sections. Drawings of plants, some of which are realistic and others which seem somewhat fantastical, make up the majority of the manuscript. Illustrations of the planets and the zodiac are in the second part. Thirdly, there is a section identified by pictures of naked ladies submerged in water. Drawings of pharmaceutical bottles and plant roots identify the fourth section, and the final section has stars on the left edge which identify it [2]. Each part is given the following names: botany, zodiac/astronomy, balneology, recipes, and starred paragraphs, respectively [1].

2.2. Scribes and Hands

Manuscript, as its name suggests, is a Latin term for a text that has been written by hand. The question is, how many hands? The majority of the information in early research on this problem is the result of supposition rather than genuine findings using the scientific method [7]. According to Albert Howard Carter, there are no discernible changes in the handwriting on either the stroke or line level across the entire text, making it very consistent and therefore created by a single scribe [8]. Erwin Panofsky, a specialist in the Middle Ages and the Renaissance, conducted further investigation of the manuscript in 1931. He later commented on it in 1954, stating that, while there is no way to be sure, the book appears to be the product of one individual, save for the last page [9].

The idea that the text in the Voynich manuscript was written by more than one scribe was first proposed by Captain Prescott H. Currier. He said that there were two degrees of distinction between the two scribes he found, the first being the handwriting variations that could be seen and the second being statistical variations in pattern at the character level [10]. These are referred to by him as several "hands" and "languages," respectively. When referring to "languages," he is referring to systems for encoding natural languages, with the key distinctions being word frequency and character sequences that are common in one language but not the other.

Statistics on the number of monographs (single letters), digraphs, and trigraphs, as well as where in a word they appear, led to Currier's classification of languages A and B. D'Imperio used cluster analysis techniques to test Currier's findings by counting the first 350 to 400 characters of 40 different manuscript pages using a monographic frequency count, taking into account that Currier believed each page to only have one "hand" and "language" [21]. As recently as 2020, Dr. Lisa Fagin Davis examined the writing of the book from a digital palaeographic perspective. This investigation used a combination of conventional approaches and technological tools, including the Mirador shared-canvas viewer, the Archetype (DigiPal) application, and the VisColl application. These resources were used for the annotation and analysis of characters. Her research led to the discovery of five distinct hands as the scribes behind the entire work [1].

2.3. Authorship Attribution

It is important to note the difference between an author and a scribe, where a scribe is the one who physically writes the words, the author is the one who dictates them. It could be the case that these two roles are filled by the same person, or that scribes are at liberty with how to write the author's ideas. This second point will be assumed to be the case in this paper, along with the assumptions that the scribes of the Voynich manuscript are multiple and are not the same person(s) as the author/s.

It is possible to infer, confirm, or characterise the author of a document in the field of study known as authorship attribution, from which inspiration for this paper's experiments was drawn. Its value has been demonstrated in a variety of situations, such as when establishing the authenticity of a document's claimed author or, conversely, when revealing an imposter. Linguistically, authorship attribution considers syntactic, distributional, frequency, and part-of-speech statistics of vocabulary as well as other properties as features in determining authorship.

As aforementioned, an initial selection of a set of features that are available and relevant in the particular case, i.e. features that will effectively indicate the authorship, is required for the successful deployment of an authorship attribution system. Character n-grams were used in numerous studies which mention their usefulness as features [12]. Similar traits in many manuscripts suggest shared authorship, whereas dissimilar features suggest the contrary. Then, a method or many methods of categorization are considered, choosing the one that produces the best results in a similar situation [11].

Support Vector Machines is a reasonable algorithm to use due to their superior capacity to process a wide range of features [13]. For accurate authorship verification, the Multinomial Naive Bayes classifier has also been combined with linguistic features successfully [12][14]. Research published in 2014 [15] develops and evaluates a neural network for language identification, which is a highly similar problem to that of authorship attribution. *A Survey on Stylometric Text Features* describes multiple instances where neural networks are successfully applied in authorship identification settings [12]. It has been demonstrated that the ReLU activation function for the hidden layers of neural networks is well suited for tasks like sentiment analysis, which is text-based and has a significant amount of data sparsity [16]. In the study *Applying Authorship Analysis to Extremist-Group Web Forum Messages* [17], a Decision Tree and a combination of structural, syntactic, and lexical variables were utilised to create a model that had an authorship attribution accuracy of 90%.

3. Methodology

The methodology described in this section is designed to test the hypothesis that Davis's proposed palaeographic classification of the Voynich manuscript into scribes [1] can cross-validate. In order to accomplish this, the manuscript's text is first collected as computer readable data. A model can only be trained using data from the Voynich manuscript because no other manuscripts utilising this particular alphabet are known to exist. To classify the entire dataset without using any external data, K-fold cross-validation is used. In every fold, a held-out sample of 10% of the manuscript is re-classified using models trained on Davis's classification for the other 90%. The data is divided into 10 folds, each of which is classified independently using a model that was trained on the other 9 parts. Four distinct classifiers were trained using 10-fold cross-validation, as shown further on in Figure 1.

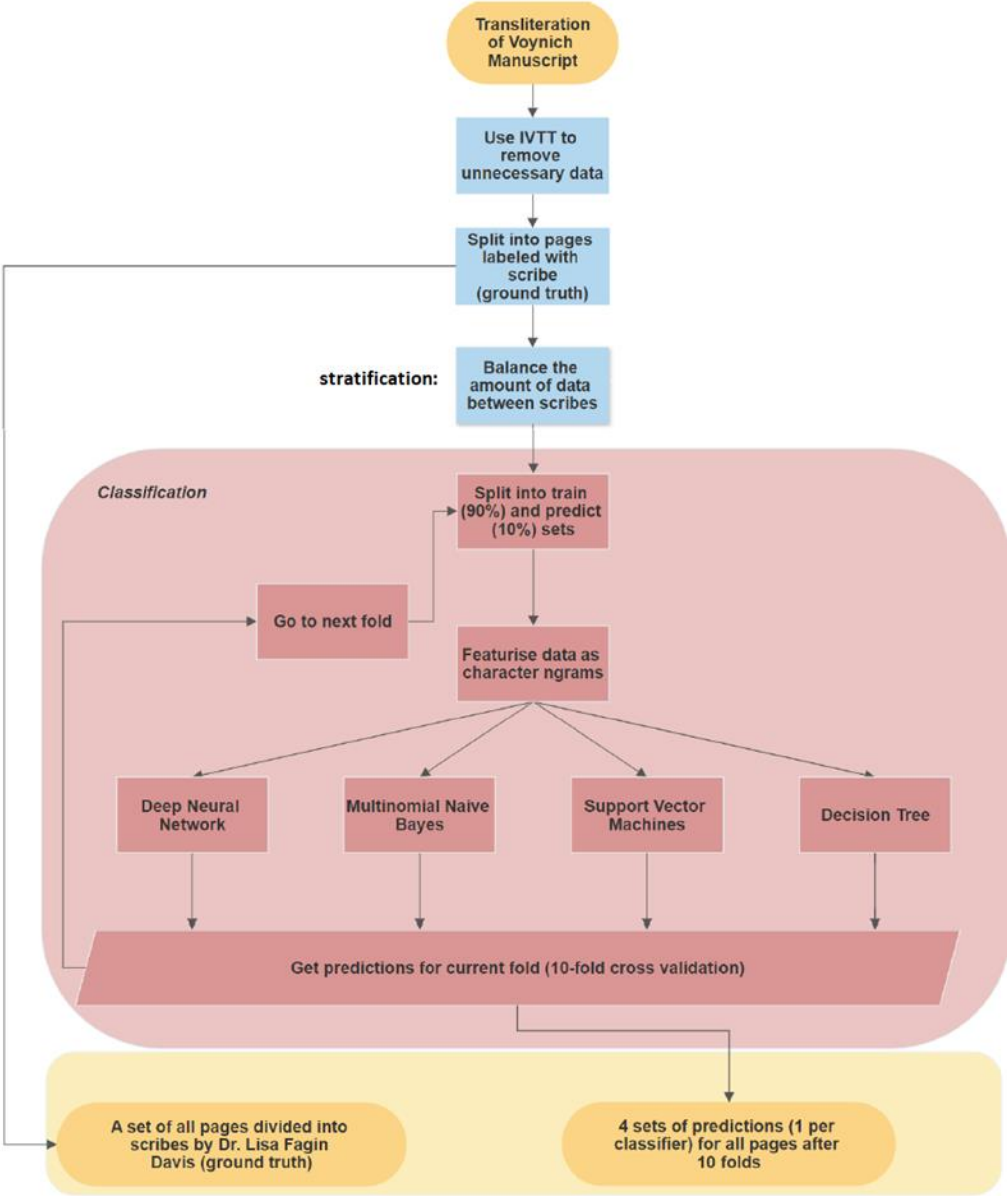

Transliteration of Voynich Manuscript
Use IVTT to remove unnecessary data
Split into pages labeled with scribe (ground truth)
stratification:
Balance the amount of data between scribes
Classification
Split into train (90%) and predict (10%) sets
Go to next fold
Featurise data as character ngrams
Deep Neural Network
Multinomial Naive Bayes
Support Vector Machines
Decision Tree
Get predictions for current fold (10-fold cross validation)
A set of all pages divided into scribes by Dr. Lisa Fagin Davis (ground truth)
4 sets of predictions (1 per classifier) for all pages after 10 folds

Figure 1: Flowchart representing the experiment's setup

3.1. Data Processing and Analysis

A machine-readable transliteration will serve as the basis for carrying out the intended experiment on the Voynich Manuscript. The transliteration utilised in this study is written with the Extensible Voynich Alphabet (EVA), one of several available transliterations written in several alphabets created especially for the task. Since it is subject to modifications, version 1r of 11/04/2020 by René Zandbergen, was obtained and used as the dataset. This file is compatible with the Intermediate Voynich Transliteration File Format version 1.7. (IVTFF) [18].

Numerous facets of the document, many of which are outside the purview of this study, are covered in the file format. The dataset had to be cleared of such extra information as the first step. Here, the Intermediate Voynich Transliteration Tool (IVTT) [19] served as a prime example of the advantages of selecting a file that supports the IVTFF format. It is a tool designed to read and process files in the IVTFF format, and its primary use is processing transliteration files to add or remove specific details about the manuscript or specific sections of the text. The decision to use that file as the transliteration file was made based on the availability of such a program. All non-EVA text is removed, except for the information on the organisation of text in the manuscript. Further data cleaning was done using Python such as ensuring that there are no duplicate whitespaces in between words.

Table 1

Page, word, and character frequency per scribe. Characters include whitespace.

Scribe	Page Frequency	Word Frequency	Character Frequency
1	107	10,005	61,016
2	42	10,218	64,295
3	31	10,698	70,806
4	15	3,155	20,377
5	7	829	5,309

As can be seen in Table 1, scribes 1 through 3 have around the same number of words and characters, however scribes 4 and 5 have relatively low counts for both. For the experiment to proceed, these numbers needed to be balanced, and the best method to achieve this is to eliminate text from each scribe until they all had data that was roughly the same size, with scribe 5 being the lowest. Unfortunately, this would produce a data set that is small, which limits the effectiveness of any machine learning algorithm. The creation of data to increase the word and character counts for scribes 4 and 5 is an additional option but generating such data from the Voynich manuscript as the sole source for these small classes would generate bias. For these reasons, it was decided to concentrate just on the first three scribes and drop scribes 4 and 5 from the experiment.

We decided to balance the number of pages per class to avoid bias. By equating the number of pages taken from each scribe class to train the models, we obtained balanced classes. We decreased both scribes 1 and 2's quantity of pages to 31, the lowest of the three, as can be seen in Table 1.

Table 2

Page, word, and character frequency per scribe after page balancing.

Scribe	Page Frequency	Word Frequency	Character Frequency	Characters per word
1	31	4,660	28,643	≈6.147
2	31	9,241	58,366	≈6.316
3	31	10,698	70,806	≈6.619

Text data on each page may vary due to illustrations taking up space, rather than text. Hence the discrepancy in page to word ratios in Table 2. When looking at the manuscript itself this becomes clearer, as the pages attributed to scribe 1 are much less text-dense than those attributed to the other two scribes. The word-to-character proportions of each scribe show that scribe 1 tends to use shorter words and scribe 3 the longest.

3.2. Classification

With regards to cross validation, previous work [20] recommends 10 folds. To accommodate the prior balancing attempts, the data is initially shuffled; otherwise, there is the danger of having a test set that contains only data from one scribe. As a trained model must accurately predict data that it has never seen before, the data is divided into 10 batches by the k-fold algorithm, of which one will serve as a test set and the other nine will be pooled to serve as the training set. To prevent data leakage, feature extraction always takes place after data splitting. Based on previous work, we decided to use character bigrams (n=2) and trigrams (n=3) to extract features in two independent runs of the experiment, to answer the second research question of this paper (section 1). When compared to word n-grams, using low-value character n-grams significantly reduces dimensionality.

The classifiers can avoid constructing a model that recognises topics rather than the stylometric characteristics of scribes thanks to this restriction. On each k-fold iteration, the high frequency (assumed cut-off of 50) n-grams are selected, creating a set of all distinct n-grams from all scribes; containing the features needed as the vocabulary for the feature matrix. The text is initially converted into a matrix of n-gram counts. The counts are then transformed using term frequency-inverse document frequency (TF-IDF). Both a training feature matrix and a testing feature matrix are created. Another data leakage preventative measure is using the training feature matrix's minimum and maximum values after min-max normalisation, as well as the vocabulary, for normalising the testing matrix.

A classification algorithm utilises labelled data as input to teach itself how to classify unseen data. Different algorithms employ a variety of techniques to accomplish the same task. For this experiment, four algorithms—a Deep Neural Network classifier, a Multinomial Naive Bayes classifier, a Support Vector Machine classifier, and a Decision Tree classifier—are taken into consideration to produce a range of results.

A neural network with numerous hidden layers is referred to as a deep neural network. There must always be an input layer, at least one hidden layer, and an output layer in a neural network. Perceptron neurons, which use weights to calculate outputs based on numerical input, are present in the hidden layers. The activation function is used to transform the neurons' output into the layer's final output. The rectified linear unit activation function (ReLU) is selected for the experiment's input and hidden layers. Due to having multi-class predictions, a SoftMax function in the output layer converts all input values into probabilities by converting them to values between 0 and 1. The target variables must be converted into one-hot encodings, which calls for the creation of a binary variable for each distinct target variable, in order to employ this activation function. The number of neurons in the input layer is equal to the number of features in the data, and the number of neurons in the output layer is equal to the number of class labels—these being the three scribes that a page can be categorised as. The batch size is set to 32 and the number of epochs is set at 5.

The same approach is used to implement Multinomial Naive Bayes, Support Vector Machines, and Decision Tree Classifiers. Predictions are then made using the test feature matrix after the models have been constructed using the training feature matrix. For issues with multiclass text classification, multinomial Naive Bayes is widely utilised. To categorise the feature space, support vector machines compute a hyperplane with a maximal margin width. The Decision Tree classifier creates a model that can predict the value of a target variable by deriving decision rules from the characteristics. The folder containing the data, code and full results of the experiments can be accessed with this link: https://drive.google.com/drive/folders/1extY_zleORyb1t4sYY9WA7bzchlCH5d6?usp=sharing.

4. Results and Discussion

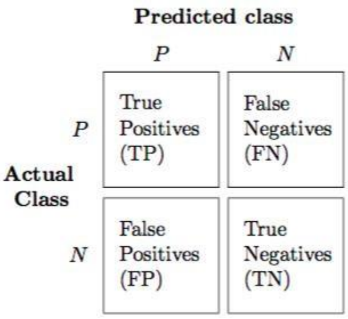
Let us start by inspecting the top-5 most frequent trigrams and bigrams. Comparable similarities may be found between the findings when analysing the top five n-grams. As seen in Table 3, the two most frequent n-grams share "y." indicating that this is the Voynich manuscript's most frequent manner of ending words.

Table 3
The top 5 most frequent trigrams and bigrams and their corresponding counts.
Note that "." represents whitespace.

Trigram	Count	Bigram	Count
Dy.	5,202	y.	10,801
In.	4,227	Ch	7,060
.qo	4,221	He	6,353
.ch	3,824	.o	5,705
che	3,821	dy	5,378

About half of these words have a "d" before "y" character. By increasing the number of characters in the sequence by 1, there was a noticeable change in dimensionality. Prior to feature extraction, there were 2,125 unique trigrams in the full dataset used for the experiment, of which 329 had a frequency of 50 or higher. On the other hand, 119 out of 437 distinct bigrams had the necessary frequency.

Below one can see the confusion matrices for each classifier following a 10-fold cross-validation. A confusion matrix reveals if a classifier consistently agrees with the ground truth and, if not, which other class was predicted in its place. The rows in each table indicate the ground truth and should all add up to 31. The columns in each table show the classifier's predictions. The number of times the classifier agrees with the ground truth is represented by the top-left to bottom-right diagonal of numbers, while the remaining values represent the number of discrepancies. Darker coloration matches higher in a coloured heatmap. confusion matrix can be represented in the manner as seen in Figure 2.



Predicted class

		P	*N*
Actual Class	*P*	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Figure 2: Confusion matrix

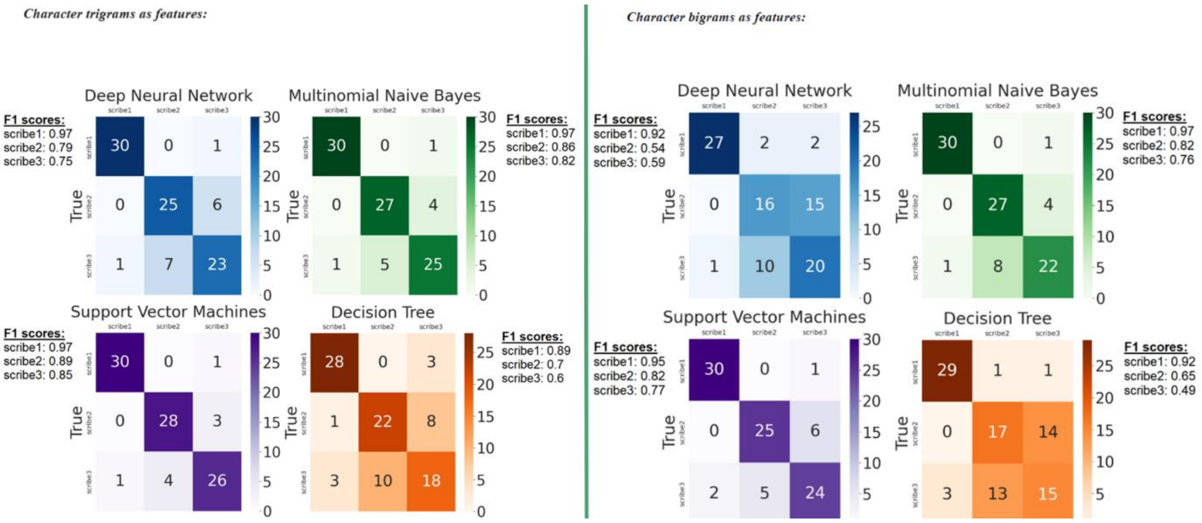
The accuracy of the classifiers will be the factor in this experiment that is given the most weight. The formula for calculating accuracy is

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{1}$$

where we divide the total number of predictions that match the ground truth by the total number of all predictions made.

Table 4
Accuracies across classifiers.

Classifier:	Deep Neural Network	Multinomial Naïve Bayes	Support Vector Machines	Decision Tree
Character trigrams:	83.9%	88.2%	90.3%	73.1%
Character bigrams:	67.7%	84.9%	84.9%	65.6%



Character trigrams as features:

Deep Neural Network

True	scribe1	scribe2	scribe3
scribe1	30	0	1
scribe2	0	25	6
scribe3	1	7	23

F1 scores:
scribe1: 0.97
scribe2: 0.79
scribe3: 0.75

Multinomial Naive Bayes

True	scribe1	scribe2	scribe3
scribe1	30	0	1
scribe2	0	27	4
scribe3	1	5	25

F1 scores:
scribe1: 0.97
scribe2: 0.86
scribe3: 0.82

Support Vector Machines

True	scribe1	scribe2	scribe3
scribe1	30	0	1
scribe2	0	28	3
scribe3	1	4	26

F1 scores:
scribe1: 0.97
scribe2: 0.89
scribe3: 0.85

Decision Tree

True	scribe1	scribe2	scribe3
scribe1	28	0	3
scribe2	1	22	8
scribe3	3	10	18

F1 scores:
scribe1: 0.89
scribe2: 0.7
scribe3: 0.6

Character bigrams as features:

Deep Neural Network

True	scribe1	scribe2	scribe3
scribe1	27	2	2
scribe2	0	16	15
scribe3	1	10	20

F1 scores:
scribe1: 0.92
scribe2: 0.54
scribe3: 0.59

Multinomial Naive Bayes

True	scribe1	scribe2	scribe3
scribe1	30	0	1
scribe2	0	27	4
scribe3	1	8	22

F1 scores:
scribe1: 0.97
scribe2: 0.82
scribe3: 0.76

Support Vector Machines

True	scribe1	scribe2	scribe3
scribe1	30	0	1
scribe2	0	25	6
scribe3	2	5	24

F1 scores:
scribe1: 0.95
scribe2: 0.82
scribe3: 0.77

Decision Tree

True	scribe1	scribe2	scribe3
scribe1	29	1	1
scribe2	0	17	14
scribe3	3	13	15

F1 scores:
scribe1: 0.92
scribe2: 0.65
scribe3: 0.49

Figure 3: Confusion matrices for both experiments

Using the confusion matrix one can calculate precision, recall and F1 score. Precision is the ratio of predictions for a class that agree with the ground truth over the total number of times that the same class was predicted. Recall, also known as sensitivity, is the number of predictions for a class that agree with the ground truth over the total number of ground truth values for that class. The F1 score is the weighted average of the previous two metrics. All classifiers in both experiments with trigrams and bigrams had high F1 scores for scribe 1 (see Figure 3. NOTE: an enlarged version is available with the results in the previously provided link in section 3.2). When compared to the other scribes, it is always the highest. This information is represented in the confusion matrices, where one can see using the heatmap that the other two scribes are rarely assigned to pages that are labelled as scribe 1 in the ground truth. The classifiers frequently classify papers marked as scribe 2 as scribe 3, and vice versa. We can infer from these results that the pages classified as being from scribe 1 are more consistent from a character n-gram perspective than the pages attributed to scribes 2 and 3, because the machine learning experiments show the greater difficulty these algorithms have with learning a good model for scribes 2 and 3 than they have for scribe 1.The difference between the accuracy scores (Table 4) of the classifiers between the two tests demonstrates that this is more common during the experiment with character bigrams. These results are also supported by the ensemble percentage agreement scores. Using bigrams resulted in a little more than half of the pages being classified by all classifiers in the same way as in the ground truth. This value jumps up to nearly 70% when using trigrams. The root nodes of both decision trees that were generated have characters in common - with the trigram being ‘edy’ and the bigram being ‘ed’. This finding matches with the fact that ‘edy’ was one of the main features that separated Currier’s Language A and Language B [22].

With regards to research question 3 of this paper (section 1), when character trigrams were used in the experiment, there were seven cases in which all four classifiers disagreed with the ground truth and agreed with one another: f57v, f65v, f85r, f86v, f95r, f95v and f116v. This happened six times when bigrams were used: f39v, f65v, f85r, f86v, f95r, f95v. In the experiment utilising bigrams, 14% of all pages were categorised as being different from the truth by the majority of classifiers, compared to 10.8% in the trial using trigrams. When at least half the classifiers disagree, these values rise to 22.6% and 14%, respectively. The fact that some pages, like folio 116 verso, have few to no words could be explained by the fact that a sizable image occupies the most of such pages. Despite the precautions made to prevent such pages from being included in the dataset, it was not possible to entirely remove all pages with low word counts without leaving an insufficient amount of page-split data to work with.

All in all, we can conclude that there is a reasonable overlap between the classifier predictions and the ground truth taken from palaeographic work. We have therefore found additional evidence for the division in scribes proposed by Davis [1] using automatic methods from the area of authorship attribution. The few examples where all four classifiers agreed on a different attribution, than the one proposed based on palaeographic work may be interesting examples for further palaeographic and stylometric study.

5. Conclusion

The outcomes of the experiments can be used to directly deduce the main answers to the research question of this paper. The results from using machine learning methods based on character sequences seem to corroborate Davis’s claims [1] that are based on the palaeographic nature of the pages. We have thus found additional evidence for the separation of scribes as proposed in earlier work with a distinct method. This opens the door for additional study in order to conduct a thorough examination of every folio of the Voynich manuscript. An alternative to this study could be a change in the features used to identify the scribes. There is plenty of literature that highlights a number of possibilities that have not yet been explored in the context of the Voynich manuscript but are used in broader solutions to the authorship attribution problem. Namely, making use of sequential pattern mining is a viable option to extract features when word meanings are yet unknown. Attention could be given to pages or folios where a majority of, or all, classifiers agreed to disagree with Davis.

6. References

[1] Davis, L. F. “How Many Glyphs and How Many Scribes? Digital Paleography and the Voynich Manuscript.” Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies (2020),5(1), 164–180. https://doi.org/10.1353/mns.2020.0011

[2] Bowern, C. L., Lindemann, L. “The Linguistics of the Voynich Manuscript.” Annual Review of Linguistics (2021), 7(1), 285–308. https://doi.org/10.1146/annurev-linguistics-011619-030613

[3] Barabe, J. G. “Materials Analysis of the Voynich Manuscript.” Beinecke Rare Book and ManuscriptLibrary (2009, April 1).
URL: https://beinecke.library.yale.edu/sites/default/files/files/voynich_analysis.pdf

[4] Zandbergen, R. “Voynich MS - Description of the MS.” The Voynich Manuscript (2018, February 8). URL: http://www.voynich.nu/descr.html

[5] Layfield, C., van der Plas, L., Rosner, M., Abela, J. “Word Probability Findings in the Voynich Manuscript.” Proceedings of 1st Workshop on Language Technologies for Historical and Ancient Languages. Language Resources and Evaluation Conference (LREC 2020), Marseille, France.

[6] Montemurro, M. A., Zanette, D. H. (2013). “Keywords and Co-Occurrence Patterns in the VoynichManuscript: An Information-Theoretic Analysis.” PLoS ONE (2013), 8(6), e66344. https://doi.org/10.1371/journal.pone.0066344

[7] D’Imperio, M. E. “The Voynich Manuscript - An Elegant Enigma.” Books Express Publishing (1978)

[8] Carter, A. H. “Some impressions of the Voynich Manuscript.” The Voynich Manuscript (1946, September10). URL: https://voynich.net/reeds/docs/carter.txt

[9] Zandbergen, R. “Special Topics: History of MS Research.” The Voynich Manuscript (2020, December13). URL: http://www.voynich.nu/extra/sp_solvers.html#panofsky

[10] Zandbergen, R. “Currier papers.” The Voynich Manuscript (2015, June 5). URL: http://www.voynich.nu/extra/curr_main.html

[11] Juola, P. “Authorship Attribution.” Foundations and Trends® in Information Retrieval (2007), 1(3), 233–334. https://doi.org/10.1561/1500000005

[12] Lagutina, K., Lagutina, N., Boychuk, E., Vorontsova, I., Shliakhtina, E., Belyaeva, O., Paramonov, I., Demidov, P. “A Survey on Stylometric Text Features.” 25th Conference of Open Innovations Association(FRUCT) (2019). Published. https://doi.org/10.23919/fruct48121.2019.8981504

[13] Diederich, J., Kindermann, J., Leopold, E. “Authorship Attribution with Support Vector Machines.” Applied Intelligence (2003), 19(1/2), 109–123. https://doi.org/10.1023/a:1023824908771

[14] Bhanu Prasad, A., Rajeswari, S., Venkannababu, A., Raghunadha Reddy, T. “Author Verification Using Rich Set of Linguistic Features.” Information and Decision Sciences. Advances in Intelligent Systems and Computing (2018) Vol. 701, 197–203. Springer.

[15] Simoes, A., Almeida, J. J., Byers, S. D. “Language Identification: a Neural Network Approach.” 3rd Symposium on Languages, Applications and Technologies (2014) Vol. 38, 251–265. Leibniz-Zentrum für Informatik.

[16] Glorot, X., Bordes, A., Bengio, Y. “Deep Sparse Rectifier Neural Network.” Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (2011), 315–323.

[17] Abbasi, A., Hsinchun Chen. “Applying Authorship Analysis to Extremist-Group Web Forum Messages.” IEEE Intelligent Systems (2005), 20(5), 67–75. https://doi.org/10.1109/mis.2005.81

[18] Zandbergen, R. “IVTFF – Intermediate Voynich MS Transliteration File Format.” Issue 1.7 (2020)

[19] Zandbergen, R. “IVTT – Intermediate Voynich MS”Transliteration Tool – User Manual.” Issue 1.1 (2020)

[20] Kohavi, R. “A Study of CrossValidation and Bootstrap for Accuracy Estimation and Model Selection.” International Joint Conference on Artificial Intelligence. (1995)

[21] D’Imperio, M.E. “An Application of Cluster Analysis to the Question of ‘Hands’ and ‘Languages’ in the Voynich Manuscript.” Unclassified Document No. 6588666, National Security Agency

[22] Zandbergen, R. “Currier A and B: two different languages?” The Voynich Manuscript (2019, June 22). URL: http://www.voynich.nu/extra/lang.html