Thesis

A Thesis
Presented to
The Division of Mathematics and Natural Sciences
Reed College

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Arts

Emily Palmer

May 2018

Approved for the Division
(Mathematics)

Andrew Bray

Acknowledgements

I want to thank a few people.

Preface

This is an example of a thesis setup to use the reed thesis document class (for LaTeX) and the R bookdown package, in general.

Table of Contents

List of Tables

List of Figures

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Dedication

You can have a dedication here if you wish.

Chapter 1

1.1 Introduction

In everyday conversation and music liturature, music is often often described using subjective summary "statistics". For example we talk about Beethoven's Symphony 9 (think ode to joy) as 'majestic', 'powerful', 'expertly written', 'etc', music critics go more in depth to talk about the performance and reception, and those with a musical background might go into more description into the actual music, couterpoint, melodic arc, etc., compared to the emotional affect.

What if one wanted to compare Beethoven to Bach? One might say that Beethoven was a classical composer, whereas Bach was a baroque composer. What exactly to those classifiers mean? Sure, it is very well documented the years each composer was active in, and that there were large changes in the popular aspects of classical music. Even the untrained ear can distinguish differences between Bach and Beethoven. What exactly is the difference that one can hear? Does Bach follow counterpoint rules more exactly? What ways can we empirically differentiate those two composers? How about contemporary composers such as Mozart and Saliery? While those familiar with classical music can spot the differences between these two composers, it is more difficult for the untrained ear or eye to spot the differences.

What if we had a piece and didn't know who wrote it? What features are inherent in a composers style that would allow us to identify the true composer?

1.2 Literature Review

1.2.1 Federalist Papers

Text classification has recently become a large field. One of the earliest instances of text classification was on the Federalist papers.(Mosteller & Wallace, 1964). The famous Federalist Papers were written under the pen name 'Publius'. There are several disputed papers attributed to James Madison or Alexander Hamilton. Historians have often examined the papers using styles of previously known writings of Madison and Hamilton. () Using the frequency of words such as and 'by', 'from', and 'upon' Mosteller and Wallace trained the writings on a set of pieces of each. These unconscious indicators were able to differentiate between the two writers, and when a model was trained (using ....), the model was able to identify the author of the disputed paper.

1.2.2 Music

Almost any piece by any composer has already been thoroughly examined by music historians. (Paragraph about things that music historians talk about when analyzing music. find sources for this.)

The human eye, no matter how well trained in music, has an extremely hard time noticing small features throughout a piece. Even if one has a feature in mind, would one want to count the number of times an author used the word ‘as’ in a 500 page book? Would one trust that count to be accurate? Say one composer was very fond of middle C, and consistently used it slightly more than other composers. Unless the use of C was extreme, likely breaking rules of counterpoint and making odd melodic and harmonic choices, a human might not be able to catch this characteristic. Writing has certain rules of grammar, that one would expect all published writers to mostly follow. Writers would likely never have the word ‘as’ written twice in a row. Similarly, classical music has rules and conventions. Counterpoint (described in Chapter 2), melody, and characteristics of the instrument composed for constrict a composer.

How can one find similar unconscious features comparable to word frequency for a composer? There are two main ways to look at features in a piece of music. The first option is using a recording of the piece. This way has been used succesfully using a perfomrance in MIDI format to distinguish certain genres of music. (De León & Inesta, 2003) They used self-organizing neural maps to classify music as either jazz or classical.

The second is using the sheet music of the piece. Performances of the same piece vary extremely, and there is information at a higher resonlution with easier access (intervals, etc.) in sheet music. Thus, in this paper sheet music will be used for analysis.

Text analysis, such as in the federalist papers, is read one word after another. Information in piece of music, however, is read in a variety of ways. It can be read left to right note by note, but it can also be read vertically as the harmony, or the notes played together. Also in a piece with several instruments, the above happens at the same time for each instrument. There are also aspects that take place over large sections, such as phrasing, or cadencial patterns. There are rules of counterpoint that are followed throughout the entire piece. Thus we need to find ‘features’ or ‘variables’ (see below) that can be measured for each piece, or perhaps each measure or instrument, that can describe a certain piece of music. Then we must decide which features are those of rules and practices of classical music, and where the creativeness and individuality of a composer happens?

Most of the musical stylometry papers have focused on composers in the Renaissance, Baroque, and Classical eras. The Mendelssohns were composing in the Romantic period. This choice might be because composers in earlier eras had less “expressive” allowances for their composing, thus making features easier, although this is just speculation. There are also more pieces with doubtful authorship in those eras.

1.2.3 Classification

Assigning likely composers to a piece of music is a classification problem. Classic approaches to this problem are described in the next section.

In our case, for each song, we will have vector X which will be a vector of all the p variables and features we are measuring. These will be the predictors. For each song we will also have, or be trying to predict y, or the identity of the composer.

1.2.4 Background on Variable and Feature Selection

Especially in research regarding gene expression and text categorization, data sets have enormous numbers of variables. We use "feature" and "variable" interchangably, with the exception when features are created from variables, and the distinction will be made in that case. (Guyon & Elisseeff, 2003). There are several variable selection algorithms that select the "important" variables. If we included every variable that we extract from the piece, our model would very likely be overfit.

The start of feature selection is domain knowledge. Thank to John Cox in the music department for suggesting a list of valuable features. These will be described in Chapter 2.

Several variable selection algorithms include variable ranking. Variable ranking uses a score function to assign a score to each possible variable. It is a computationally efficient method and is robust against overfitting as it introduces bias but may result in less variance. It is tempting to only include variables that have a high score. However, this possibly leads to redundancy. In addition, variables that are not important by themselves can have a significant performance improvement when considered with other variables. Popular variable ranking methods for classification are single variable classfiers and information theoretic ranking criteria.

Single variable classifiers rank the variable according to thier individual predictive power. The predictive power can be measured in terms or error rate, or using the false positive or false negative rate (fpr, fnr). This classifier cannot distinguish variables that perfectly seperate the data.

The Information Theoretic Ranking Criteria is used in variable selection. They often rely on estimates of the mutual information between the predictor and response, as given by

$$I(i) = \int_{x_i} \int_y p(x_i, y) \log \frac{p(x_i, y)}{p(x_i)p(y)} dxdy$$

where $p(x_i)$ and $p(y)$ are the probability densiites of x_i the $i^t h$ predictor and y the response, and $p(x_i, y)$ is the joint density. $I(i)$ is a measure of dependency between the density of variable x_i and the density of the response y (reword)

After knowing the ranking of a variable we then select which variables will be useful for our model. This is known as variable subset selection. The three most common types of variable subset selection are wrappers, filters, and embedded methods. Filters do not involve any machine learning to create the criterian for varable subset selection. Wrappers on the other hand use the "performance of a learning machine trained using a given feature subset." Embedded methods perform variable selection in the process of training and are usually specific to given learning machines

All possible subsets of variables is $2^p - 1$, which for large p is often computationally impossible. Strategies like best-first, branch and bound, simulated anealing, and genetic algorithms can help with the computational difficulties.

Wrappers are often thought of as brute force methods. This can be good, as it can reduce overfitting. Two types include forward selection and backward elimination. These both give nested subsets of variables

Often there is a need for dimensionality reduction. Is there a way to combine enough of the information given in the features in a smaller dimensinoal space? This results in feature creationg; using the recorded variables to create new features to fit the model on.

These include clustering, basic linear tranformations of the input variables, such as PCa/SVd, and LDA. Also more sophistecated linear transformas like Fourier and Hadamard.

Two basic goals of these feature creations, are that we can achieve a good reconstruction of the data. The second is that we can be most efficeint in making our predictors. The first is an unsupervised problem. The second is supervised.

Clustering is in fact a type of feature construciton. The group of clusterd points thereby becomes a feature. Examples of this include K-means and hierarchical clustering.

SVD, singlar value decomposition is another form of feature construciton.

Trees

K-means

PCA

Floating Forward Selection

1.2.5 Previous choices of features

Deciding on and extracting features of music is the first step to analysis. Depending on the characteristics of the composer and time period, different features would be useful. Often, features are extracted en masse and then work is done later to determine which features are important or useful in identifying style.

In addition to what kinds of features, in music there is also the question of the scale at where those features take place. They can be features for a given instruemnt, the entire piece, or each measure. Also windowing techniques can be used where a "window" is created over a given number of bars or notes, and moves thorugh the whole piece. For each window, a feature is recorded. These can be overlapping windows by creating an "ofset" of a number of beats or notes. This produces more data, as instead of one feature for each piece, there is a feature for each window, and there an be tens of windows in each piece.

Common types of features used before in music analysis are: Frequencies or fractions of notes, chords, etc are a common low-level feature. These include the fraction of the score that consisted of dissonant sororities, as well as the fraction of bars that begin with a dissonant sonority

Features measuring "stability" are also popular. Stability is computed by dividing the standard deviation of the lengths of the fragment by the mean length of the fragments. It is normalized in this way to be comparable over differing time signatures. (Backer & Kranenburg, 2005)

Markov transition matrices for the rhythms of the pieces, and Markov transition matrices of the pitches in each piece

The above teqniques were used to analyze the music of Bach, Handel, Telemann, Mozart and Haydn and compare J.S. Bach, W.F. Bach and J.L Krebs in an attempt to classify BWV 534.(Backer & Kranenburg, 2005) They use overlapping windowing over each entire composition to produce more data, and avoid issues of dimensionality. They chose a window of 30 bars to create a high enough number of fragments per piece and a low enough variance of the feature values between fragments. They chose to extract 20 features including features of freactions and measuring stability, and entropy.

Aditionally, a number of previous papers have focused on Josquin des Prez. This is likely due to the fact that there is a large training and testing data set available in easily analyzable format provided by the Josquin Research Project (citation). In addition there are a number of pieces of disputed authorship that have been attributed to him. Work by Brinkman et al. (Brinkman, Shanahan, & Sapp, n.d.) use machine learning approaches to evaluate attribution of compositions by des Prez. They used both high level and low level features. The high level features were 9-8 suspensions, oblique motion, contrary motion, similar motion and parallel motion. The low level features were average melodic entropy, normalized pairwise variability index (?), and note-to-note transition probabilities.

Work by Speiser and Gupta (Speiser & Gupta, n.d.) analyzed Josquin and his contemporaries to attempt to classify unknown works. They extracted four categories of features, frequencies of individual notes, frequencies of pairwise interval combinations between each of the voices, Markov transition matrices for the rhythms of the pieces, and Markov transition matrices of the pitches in each piece. In total, this lead to a total of 3000 features.

1.2.6 Previous applications

Most of the previous research has needed to do some kind of feature selection. A lot of features are extracted as a priori we don't know which features are distinguishing.

A modification of a forward selection (Floating Forward Selection(cite)) was used to extract features in order to identify distinguishing style between Bach, Handel, Telemann, Mozart, and Haydn, and then subsequently classify the authorship of BWV 535. (Backer & Kranenburg, 2005) Each composer was compared via creating comparisons of all possible class arrangements, ie (Bach)(Handel), (Bach)(Handel,Teleman), etc. The algorithm extracted features for each class arrangement that distinguished the groups the best. A decision boundry was used for Bach and not Bach, on the features Diss Part, Par thirds, and stab time slice. A k-nearest neibors classifier was successful in comparing Bach and others as well as each individual composer. Decision trees to interpret the features used in decision making of the different class arrangements.

To determine authorship of BWV 535, they train a quadratic Bayesian classifier to distinguish J.S. Bach, W.F. Bach and J.L Krebs. They again compare every possible class arrangement as potential composers.

PCA was used to analyze the music of Bach, Handel, Telemann, Mozart and Haydn and compare J.S. Bach, W.F. Bach and J.L Krebs in an attempt to classify BWV 534.(Backer & Kranenburg, 2005) Although only two PC's accounted for most of the variance, 5 PCs were used to account for more variance. Binary comparisions were used to compare composers. This resulted in a relatively clear separation between Bach and Josquin. For Josquin and his contemporaries, the PC's do not do as well a job of separation. The results of the principal component analysis run on all the composers, were used to train a classifier on all the composers. First a k-nearest neighbor classifier was used. To accound for most of the variance, 27 PCs were used. Next they trained a support vector machine classifier with a radial kernel. Finally they used a decision tree to determine which features were important in discerning the composers.

Speiser and Gupta (Speiser & Gupta, n.d.) scored each feature by the mutual information of each features. They then chose the top 50 features and ran GDA. They then ran PCA to attempt to remove some of the dependencies associated with musical features. They first fit a Naive Bayes for classification, but it had a large training error as the independence assumption does not work well with musical data. Next they used support vector machines with a Gaussian kernel and GCA learning algorithms.

1.3 Fanny and Felix Mendelssohn

Most musical stylometry analysis focuses on music of the Renaissance and Baroque period, as there are more questions of authorship in that period. As the Romantic period is much more modern in comparison, there are many more surviving records of original manuscripts that include the composer.

Felix Mendelssohn, often considered a prodigy akin to Mozart, was a prolific composer. Before he was fourteen years old, he had already written over 100 compositions.

His lesser known sister Fanny Hensel was also a composer of incredible skill. The two were very close, for many years training and studying together. In their early education living in Berlin, Felix and Fanny received the same musical education, first piano lessons by Madam Bigot, a famous pianist esteemed by Haydn and Beethoven. Beginning in 1818, Carl Friedrich Zelter, a somewhat removed student of Bach and the most influential Berlin musician of the time, began to teach them both composition. In addition to music, the children were tutored by some of the finest scholars in Berlin in subjects such as languages, history, and drawing. Goethe himself claimed that Fanny was "as gifted as Felix". (Tillard, 1996)

As Fanny grew up, her father started implying that she should focus her energy on the domestic sphere of her life. While the fact that she never became a world famous composer and performer is often attributed to the gender politics of her time, it is also likely due to her high class. (Reich, 1991) Especially considering the anti-semetic feelings of the time, and since the family had recently converted from

Judaism to Christianity, the family did not want any other unusual characteristic such as a professional female composer to set them further apart from "polite" society.

Most of Fanny's available work are Lieder, short pieces of voice accompanied by piano. They were accepted at the time as the more feminine, domestic compositions, acceptable for women to compose. Her brother moved on to more elaborate compositions such as operas and orchestral concertos. Her father pressured Fanny to remain composing Lider. (Todd, 2003)

> "Music will perhaps become his profession, while for you it can and must only be an ornament, never the root of your being and doing. We may therefore pardon him some ambition and desire to be acknowledged in a pursuit which appears very important to him,... while it does you credit that you have always shown yourself good and sensible in these matters; ... Remain true to these sentiments and to this line of conduct; they are feminine, and only what is truly feminine is an ornament to your sex."

Throughout their lives, Felix and Fanny maintained contact through letters until Fanny's death in 1847 and Felix's death shortly thereafter. These letters contain many instances of Felix asking for advice on his compositions (include quote)

Unlike Felix who conducted and performed piano and organ in some of Berlin's most esteemed concert halls, most of Fanny's performances were private, only performed in small circles of her friends and family at intimate parties. Similarly, although she was quite a prolific composer, under recommendation of her father Abraham Mendelssohn, and to a lesser extent Felix, Fanny did not publish her work until later in her life. In 1846 after her fathers death and though her brothers disapproval, she published her first collection of Lieder. Many of Fanny's unpublished notebooks are in private collections and are inaccessible

However, it is widely speculated (known?) that some of Fanny's work was published under her brother's name, Especially three pieces each in his Op 8 and 9 lieder. Famously, when Felix met the Queen of England, she sang Felix's Lied "Italien", and Felix had to admit that in fact, it was his sister that had written it. In a letter to Felix, Fanny admits:

> "I have just recently received a letter from Vienna, which contained basically nothing but the question of whether "On Wings of Song" was by me, and that I should really send a list of things that are running about in the world disguised, it seems that they aren't clever enough themselves to separate the wheat from the chaff." (Mace, 2013)

As she never made such a list, we are left to wonder if there are any other pieces of hers that have been published under her brothers name and reputation.

This project will use Lieder of Fanny and Felix Mendelssohn. Most of the available work by Fanny are Lieder, of which Felix also composed a great deal. We will see if there is a determinable difference in style of these siblings who grew up very close and received mostly the same musical education. We will then look at the (disputed?) Op 8 and 9. Additionally, using Lieder that have been decidedly written by Felix, we will see if any other of his earlier publications could have potentially been written by Fanny.

Chapter 2

2.1 Summary of notation and vocabulary used in this paper

Vocab	description
piece	the entirety of one opus, referred to as a piece in any stage
staff	one instrument composed of one staff line

2.2 About the data and conversion process

2.2.1 Pieces used

In total there were xxx pieces from Mendelssohn, and xx from Fanny, They were...
insert picture of score

2.2.2 Optical music recognition

Sheet music as a form of data requires a lengthy process of conversion before being able to be used in any analysis. Simply scanning the scores into, say, a PDF, gives no musical semantics and can only be viewed on screen or printed on paper. Thus, the two main steps in reading in data from sheet music are: using optical music recognition software to transform physical scores into digital formats. The second step is to read the digital format in to R.

The scores used in this paper were obtained from physical copies available in the Reed music library. These scores were then scanned using software designed for optical music recognition (OMR). This method requires learning from graphical and textual information. The main things the software must pick up are the locations of bar lines, notes and rests, slurs, dynamic and tempo markings, lyrics etc. Basic optical music recognition has been around since 1966. Most commonly, the first step in optical music recognition is to remove the staff lines. The staff lines are critical, as they define the basis for the vertical definition distance of pitch, and the horizontal distance definition of rhythm. The staff gives a normalization that is helpful, essentially defining the size of what notes and Rhythm will look like. Staff removal methods include projections, histograms, run lengths, Candidates assemblage, contour tracking, and Graph path search. - which i don't understand and may not write about ... - (**???**) The next step is music symbol extraction and classification. These methods include template matching, where the object in question is compared to existing known musical symbols, simple operators, such as analysis of bounding boxes and projections, joining graphical primatives, such as combining extracted objects such as notes, note heads, and note beams to connect them in a musically correct way to form chords etc. Other methods use statistical models for analyzing musical primitives (the objects its trying to classify) such as Neural Networks, Support Vector Machines, k-Nearest Neighbor, and Hidden Markov Models.

The next step is syntactical analysis and validation. This step essentially uses defined grammars describing the organization of music notation in terms of music symbols. This makes the classification problem simpler, as there are existing rules and relationships between musical symbols.

insert picture of MuseScore score

The OMR used in this paper was PhotoScore. Photoscore outputs a musicXML file that can be read in by most music composing software, such as Sibelius, Finale or MuseScore. Muse score is a free software and was used in this paper. After being read into MuseScore, each piece was proof-read and corrected, as there were often errors in the OMR, especially in recognizing triplets.

MusicXML on its own is not conducive to converting into a data frame as representing the single note middle c looks like this:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE score-partwise PUBLIC
    "-//Recordare//DTD MusicXML 0.5 Partwise//EN"
    "http://www.musicxml.org/dtds/partwise.dtd">
<score-partwise>
  <part-list>
    <score-part id="P1">
      <part-name>Music</part-name>
    </score-part>
  </part-list>
  <part id="P1">
     <measure number="1">
      <attributes>
        <divisions>1</divisions>
        <key>
          <fifths>0</fifths>
        </key>
        <time>
          <beats>4</beats>
          <beat-type>4</beat-type>
        </time>
```

```
      <clef>
        <sign>G</sign>
        <line>2</line>
      </clef>
    </attributes>
    <note>
      <pitch>
        <step>C</step>
        <octave>4</octave>
      </pitch>
      <duration>4</duration>
      <type>whole</type>
    </note>
  </measure>
</part>
</score-partwise>
```

MusicXML is used commonly as it is conducive to representing sheet music and music notation, and it can be transferable to many different music softwares. We then need to convert into a format more easily readable into R. We do this by using Humdrum's function xml2hum that converts a musicXML file into a .krn file. This file time can be read much more easily into R. Compared to above, the code for a single middle c whole note would be :

```
**kern
*clefG2
*k[c]
*M4/4
=1-
1c/
```

The import_xml_files.sh file goes through the process of converting scores from musicXML to .krn. The files need to be separated into having a separate file for each staff, which would mean a seperate file for each instrument. In our case for each piece there are always two or three files for each piece, which are voice, piano right hand, and piano left hand. This is necessary to avoid the bugs in xml2hum that have issues when staffs don't necessarily match up as a result of the conversion process, most often by rhythm.

2.2.3 .krn to R

Once we have .krn files to represent each piece (usually 2 or 3 files), we use regular expressions to extract key information.

2.3 Raw data frame:

The data frame representing one piece: Each instrument contains the following columns: rhythm value [4], rhythm name [quarter note], note value [5], note name (octave inclusive)[cc], note name (octave exclusive)[C sharp]

Additional column values are measure.

Each row of the data frame contains one time base value. For a given piece, the time base represents the shortest note duration value. For example, if the shortest note a piece contained was a sixteenth note, the time base would be 16. Each measure then would contain 16 rows.

2.4 About the functions - Feature creation - liley some musical infromattion

Once the scores are converted into R raw data, feature creation begins. These features are mostly features suggested by John Cox.

Chapter 3

EDA

Chapter 4

4.1 Feature Selection

Next they computed the entropy of the probability of occurrence of ways of thinking about chords; chords are the same no matter what scale degree they are on, and distinguishing chords differently. Next the entropy was calculated given the probability of each pitch in the score. Entropy was calculated by $-\sum_{i=1}^{N} p_i \log p_i$ where p_i is the probability of occurrence, and N is the total number.Next they the average number of active voices at one time. This represents the voice density of the piece. Then for every interval, the duration of that interval was divided by the total duration of all intervals. Next the total duration of parallel thirds, fourths, and sixths divided by the total duration of all intervals was measured. Finally a measure of suspensions was found. (Backer)

Chapter 5

5.1 Model selection

Chapter 6

6.1 Model Fit

Chapter 7

7.1 Discussion

Conclusion

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readibility and/or setup.

In the main Rmd file

```
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis.
if(!require(devtools))
  install.packages("devtools", repos = "http://cran.rstudio.com")
if(!require(thesisdown))
  devtools::install_github("ismayc/thesisdown")
library(thesisdown)
```

In Chapter ??:

Appendix B

The Second Appendix, for Fun

References

Backer, E., & Kranenburg, P. van. (2005). On musical stylometry—a pattern recognition approach. *Pattern Recognition Letters*, *26*(3), 299–309.

Brinkman, A., Shanahan, D., & Sapp, C. (n.d.). Musical stylometry, machine learning, and attribution studies: A semi-supervised approach to the works of josquin.

De León, P. J. P., & Inesta, J. M. (2003). Feature-driven recognition of music styles, 773–781.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(Mar), 1157–1182.

Mace, A. R. (2013). *Fanny hensel, felix mendelssohn bartholdy, and the formation of the" mendelssohnian" style* (PhD thesis).

Mosteller, F., & Wallace, D. (1964). *Inference and disputed authorship: The federalist*. Addison-Wesley.

Reich, N. B. (1991). The power of class: Fanny hensel. *Mendelssohn and His World*, 86–99.

Speiser, J., & Gupta, V. (n.d.). Composer style attribution. *Project Report for CS*, *229*.

Tillard, F. (1996). *Fanny mendelssohn*. Hal Leonard Corporation.

Todd, R. L. (2003). *Mendelssohn: A life in music*. Oxford University Press.