

1. Introduction

Of all major art forms, music has historically relied most upon scientific and mathematical devices in its creation. While many other forms of art are lauded for breaking the rules, and these avant-garde approaches often find themselves at the forefront of popularity, the most praised and well-respected pieces of music always seem to find themselves firmly grounded in the formal rules of composition that have been widely accepted for centuries.

The reason behind this can be easily attributed to the notion that music is well founded in the world of mathematics, and the rules of music theory are indeed built upon it. Both the relations between pitches and durations are best defined by numbers and ratios. In fact, because of its reliance on precise measurement, music was considered until fairly recently its own branch of science [1]. This fact makes it tempting to both analyze and create music through a scientific approach, and it is indeed a venture that has been attempted many times over the course of human history, making great strides since the beginning of the digital age.

1.1 Early Exploration

The intersection of mathematics and music predates the computing age quite considerably. The topic of algorithmically composing music saw its initial explorations as early as 500 B.C. in the times of Pythagoras [2], when he developed the concept of “music of the spheres,” in which he drew some of the first significant connections between the world of music and mathematics. Of course, Pythagoras could not have known what he was pioneering would one day spawn the algorithmic composition of music, as the term ‘algorithm’ wasn’t even invented until 1120 [3]. From this point on, the world of music was situated comfortably in the middle of the mathematical spectrum, and a millennium later, Flavius Cassiodorus (ca. 485-575) described mathematics as a union of the four disciplines: arithmetic, music, geometry and astronomy [4].

At the dawn of the medieval era, composers began to formulate rules by which pitch relations and combinations were governed, laying the groundwork for music theory as a practice that would be followed and expanded upon for centuries [5]. It was in the 1700's with a game called *Musikalische Würfelspiel* [6], which translates from German to 'musical dice game,' that the rules were put to use in an algorithmic fashion. The game's most popular iteration, allegedly devised by Mozart himself, saw the user roll a pair of dice, and their composition would proceed based on the outcome being mapped to a ruleset Mozart outlined. These early experiments laid the ground work for algorithmic music to come.

1.2 The Data-Driven Intelligence Age

With the framework of algorithmic music already set centuries before, it was only natural that the concepts were brought into the world of computing as early as the 1950's, at the genesis of the information age. The most famous example from this time is Hiller and Isaacson's *Illiad Suite* [7], which used rule systems and Markov chains, a stochastic predictive system with no memory, to predict the next successive note based solely on the current note. As the work was expanded upon by colleagues and interested parties, the chains were designed to implement an n th-order technique, which allows the process to consider the last n notes, rather than only the most recent [6]. This initial work with Markov chains became the springboard of computerized algorithmic compositions.

Since this advent, the topic's exploration has increased drastically, and has branched into many different realms, with new techniques and structures being used as the basic building block of the composition process. In his book "Algorithmic Composition: Paradigms of Automated Music Generation," Gerhard Nierhaus split the topic into several distinct categories, including generative grammars, transition networks, genetic algorithms, cellular automata, artificial neural networks (ANNs) and artificial intelligence [3]. As these fields grow further apart, greater strides and achievements are being made within each.

The intersection of music and computing becomes even more pronounced when you approach the topic of data mining. Many have explored the potential of classifying music of all varieties, and results have been quite successful. Researchers Lebar, Chang & Yu [8] used classifiers to distinguish between the works of various classical composers using stylistic features as attributes. Basili, Serafini and Stellato [9] tackled the topic of popular music when they classified a dataset of music into six distinct genres based on features such as intervals, instruments used and meter changes. The basic structure of this study has been conducted by many, receiving respectable results overall.

It is important to note that this is not the first experiment that attempts to use classification techniques to create algorithmic compositions. One particular avenue in this field that has oft been explored is the use of artificial neural networks (ANNs). The basic structure of an ANN has allowed for a variety of approaches to music composition. Some experiments have used the structure to encourage the refinement of musically random melodic phrases, or to predict the melodic phrase based upon a number of starting notes. Others attempt to merge the predictive powers of the classifier to build upon another method of composition [6], much like our proposition. To our knowledge there are no experiments which attempt to use this classification technique, or any other, to inspire algorithmic composition through Cellular Automata.

1.3 Study Overview

While it is clear that the topic of music's intersection with computer science has been explored in many facets, there is still a gap when it comes to what a computer is capable of producing, and some of the most recent studies in the field of algorithmic composition are still labeled as composition inspiration software [6]. The idea of hybridizing multiple of the above concepts has therefore become attractive, in an effort to achieve the best generative characteristics from multiple approaches. For this reason, we find it worthwhile to explore new avenues, and see what kind of new directions we can bring to the topic of algorithmic composition.

It became evident during the course of our research that one such hybridization comes from the potential of using the field of data mining to inform the decisions made during certain algorithmic composition techniques. Intersecting these two concepts has the potential of creating a smarter generative process, capable of replicating nuanced differences between several different categories of music, adapting to new forms of music being introduced, and minimizing the amount of human intervention required for some techniques. One such intersection that we saw potential in was using data classification to inform a cellular automata composition system. It is under the guide of this general framework that we began our work.

2. Data

With any venture into the world of data mining, it is critical to choose the right data with which to proceed with your experiment. The topic of music presents a particular challenge in this respect, as the data at hand is not nearly as friendly for computer use as something purely numeric such as stock numbers or attendance projections may be. For this reason, a substantial amount of time needed to be dedicated to understanding the data of music, discovering what kind of characteristics are desirable to use from the data, and what kind of computer-friendly representations we have as options moving forward.

2.1 Musical Representation

In order to properly understand the data, it is important to first have a firm background in the formalities of music. For the sake of this experiment, we will be narrowing the scope of our focus entirely upon classical music, which we define as traditional Western music ranging from the Medieval era to the Modern era (not to be mistaken with the Classical era, which is a distinction within the realm of classical music). The main reason for this decision is classical music's written consistency across history [5]. Music has evolved and expanded greatly since the days of Mozart and Bach and as a result, much of what is being created today in popular music has abandoned the concept of formally creating a written representation of the music. Recent years have seen the greatest decline in non-educational

production of sheet music [10]. Luckily, classical music, by virtue of its creation for performances by individuals other than those composing, as well as its educational value, has a rich history of written representation. It is still most widely recorded in this manner today, and thus provides us with a much more stable and wide backlog for analyzation.

This backlog of written classical musical literature is comprised almost entirely within the medium of musical scores, or sheet music. Sheet music is a visual representation of music made up of symbols and words which convey all the information a performer must know to play the piece. Among other information, these symbols are capable of portraying which notes must be played at what time, the volume at which they are to be played, and in what rhythm. This manner of recording music started as early as the ancient Greek and Middle Eastern civilizations where they began using basic music symbols as written reminders. It wasn't until the 9th century that Christian Monks began recording music on sheets. From this point on, the practice exploded in popularity, and has maintained the same basic structure [10].

2.2 Digital Formats

For hundreds of years, Western music has been represented by means of these musical scores. This has been relatively unchanged because it is an ideal notation for a musician to read and perform [10]. With the advent of the digital age, the necessity for a new representation of written music was realized. This was due to the complex nature of musical scores. It is quite difficult to teach a computer to parse through the various symbols and notations of music, making the task of retrieving the data necessary for processing challenging. As a result, the computer science community was met with the challenge of creating a new representation of music that could be more easily processed for the studies to come. Though many were proposed, two have risen above the others in the world of research, MIDI and **kern musical files. Both have their own unique advantages and disadvantages.

2.2.1 MIDI

First seeing its start in 1981 [11], the Musical Instrument Digital Interface (MIDI) format is one of the most widely used digital musical formats that exist. By virtue of its creation for use with electronic synthesizers, MIDI files contain representations of the musical score that are often recorded via humans playing the score with a synthesizer, though you can also find hand compiled MIDI representations.

Over time, this format has been adapted for use in scholarly research, with many toolkits being developed, such as jSymbolic [9], to extract data from the MIDI files. Because of its widespread use for a variety of functions, the backlog of MIDI scores to be used for potential research is vast, but also unreliable. This is due to the fact that anyone with an electronic keyboard can plug it into a computer and create these files, regardless of their accuracy level. Despite this, we found throughout our survey of previous studies that MIDI is the most widely used file type in academic research concerning computer music.

2.2.2 **kern

While the MIDI format was created for a wide variety of computer music purposes, a format known as ****kern** was created with a much narrower intention. ****kern** files are musical representation files which fit within a broader syntax known as ‘Humdrum.’ Described by its creator David Huron as a “general-purpose software system intended to assist musical research” [12], the software was quite literally designed for use in projects like this. Researchers Lebar, Chang & Yu [8] used this format in similar research when attempting to classify musical scores by composer.

The Humdrum software can be split up into two separate entities: The Humdrum Syntax and the Humdrum Toolkit [12]. Humdrum Syntax is a grammar by which any file that falls under its guise must adhere to. ****kern** is a single file type under this syntax, and indeed the most widely used of them, designed to represent the core information for common Western Music. The format is capable of representing nearly every nuance found

within a musical score, down to the direction the stem of a note is facing on the page. The other half of the equation, the Humdrum Toolkit, is described by Huron as a toolbox of ‘utilities,’ with over 70 inter-related software tools, which can be used to manipulate any data that conforms to the Humdrum syntax [12]. These tools, combined with the vast number of features that can be represented using the Humdrum Syntax, make it an attractive option in the realm of data mining.

While this format offers many advantages, there are certainly drawbacks to it as well. Because of its rather limited usage (being designed specifically for research purposes), the amount of data available in this file type is sparse. There have been a number of people who have contributed a substantial number of scores encoded in `**kern` format, however the encoding process, which must be done entirely by hand, is a tedious one (though perhaps lends itself to a greater attention to detail), and there will never be a rich well of files to choose from.

Despite this deficiency, we found the format of `**kern` to be most compatible with the task at hand. The Humdrum toolkit offers us an effective way to extract any and all information about the score we may find useful, and the textual representation is also much friendlier to interpret on a visual level. With this decision, we began our work in data mining.

3. Data Mining

Data mining itself is a broad term, and is truly a confluence of many disciplines, including mathematics, computer science and statistics. The applications of this intellectually stimulating field are plentiful, diverse, and exciting for those focusing on the topic. In the scope of our study, data mining provides us with a tool to discover the defining features of music composition and preserve this information for the computer to use in its future music generation. The phrase ‘data mining’ itself defines a rather broad idea, simply described as “the process of discovering useful information in large data repositories” [13]. In the

pursuit of achieving this goal, data mining has been approached using several other distinct methodologies, such as classification, clustering and association, among others [13].

While each of these data mining methods have merit, and some may indeed be useful in future works while attempting to improve the algorithmic music composition challenge, this study has chosen to focus its attention on the topic of classification. Classification is defined as “the task of assigning objects to one of several pre-defined categories” [13]. This objective may be achieved through the use of a learning scheme that generates a set of rules or patterns by which data instances are classified into these pre-defined classes. The trained classifier is then able to predict the classes or categories based on the generated rules [14]. The predictive power of this form of data mining is one of the driving forces behind our decision to focus on classification, as a predictive rule-based system provides us a nice backbone upon which to build a music generator.

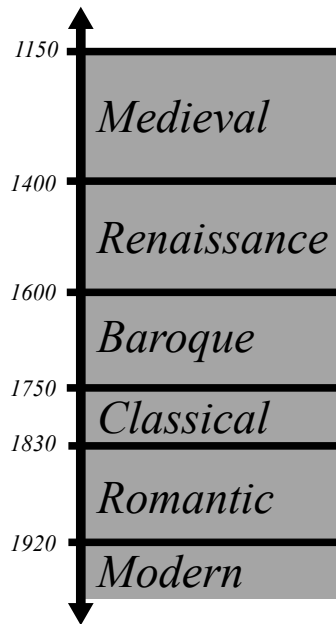
3.1 Data Extraction

In order to get the most out of the data mining process, there is a large amount of preparatory work that must be done to ensure that the information received as consequence of our work is valuable and significant. Our results are only as valuable as the system from which they were derived, so it is important to ensure we make the correct decisions leading up to the actual data mining taking place. Some of these decisions include dictating which pre-defined classes to supply our classifier, which features we would like our classifier to look at in making its categorizations, and the pre-processing and data extraction required to make the data accessible for the actual data mining process.

3.1.1 Classes

The first thing we needed to do when prepping our data for processing was select the pre-defined classes by which to separate the data, as the classification methodology necessitates. In musical classification, there have been studies that have done this in several manners, whether it be by composer, genre, or even decade. For the sake of our study, we

found it most appropriate to create the classes based upon musical era within the classical spectrum.



*Figure 1 – A timeline displaying the order and generally agreed upon
dates of the various eras of classical music*

There have been several eras by which the style of a classical piece can be defined, roughly outlined in figure 1. The years in which these eras transitioned between one another have been debated by experts [5], however it is generally accepted that there are six distinct eras, ranging from the beginnings of formally composed music in the medieval era to the wildly innovative and often atonal modern era of classical music. Moreover, students and scholars of music are able to use their training in aural skills, such as identifying the interval between any two successive notes, among other musical features, to identify which of these eras a piece of classical music belongs to. This suggests that there are quantifiable differences in their structure that make it so and provides us great reason to believe a computer will be able to identify these differences as well.

3.1.2 Attributes

Our next step was to decide which attributes we would be basing our classification upon. In data classification, these attributes – or features – are the sole factors analyzed in an attempt to generate patterns for separating the data into the pre-defined classes it has been given [13]. It is therefore important to choose features that are both indicative of the stylistic-era under which the piece was composed, as well as replicable for the future generative process. The features decided upon after consideration of a number of factors, presented in figure 2, are based upon the notion of a musical interval. The task of choosing these attributes came with two major challenges; one musical and one computational.

	Attribute	Description
X ₁	freqUni	Ratio at which unison intervals occur (unison/total)
X ₂	freqStep	Ratio at which stepwise intervals occur (step/total)
X ₃	freqThird	Ratio at which third intervals occur (third/total)
X ₄	freqFourth	Ratio at which fourth intervals occur (fourth/total)
X ₅	freqFifth	Ratio at which fifth intervals occur (fifth/total)
X ₆	freqSixth	Ratio at which sixth intervals occur (sixth/total)
X ₇	freqSeventh	Ratio at which seventh intervals occur (seventh/total)
X ₈	freqOct	Ratio at which octave intervals occur (octave/total)

Figure 2 – List and description of attributes used in classification process

By merit of the musical data we are using, there were countless numbers of attributes through which we had to sift in order to choose our features. As discussed in section 2.1, a piece of sheet music contains a vast amount of information, and our selected **kern format does little to narrow down that scope, as it does such an excellent job of preserving all the information recorded in a traditional score. Our chosen attributes must be indicative of the era the piece represents, so as to allow the classifier to accurately and practically determine which era the piece came from.

From a computational standpoint, we wanted to consider features that would lend themselves to both the classification process, as well as the generation process in the next

step of our research. Classification mandates that each feature within its system be flat rather than structural – meaning that the value can be defined by either a numeric or discrete value [14]. Because of music’s reliance on mathematics, this factor is not terribly delimiting, but it does help suggest which features may lend themselves best to the process: those which are finite and numerically categorized. It behooved us to focus on features which we could see as easily replicable in a future generative process, meaning features like dynamics, a feature that indicates how loud a particular section of the musical piece, would do little good on their own, despite being important to the construction of a musical piece.

After consideration of these factors, the decision was made to focus upon the frequency with which certain musical intervals occur within the pieces of music. Before we delve into why exactly we made this decision, it is important to understand what an interval is.

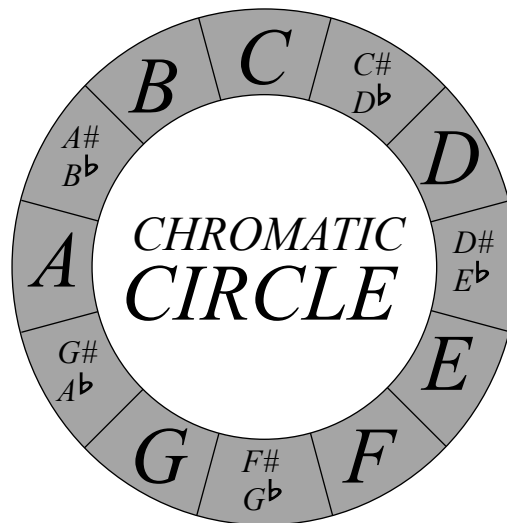


Figure 3 – A visual representation of the Chromatic

Circle, the backbone on which Western music has been created

The concept of a musical interval is built upon the very foundation of Western music: the chromatic circle (Figure 3), a cyclical scale of equal temperament made up of 12 total pitches [15]. A piece of music is comprised of a finite number of these 12 pitches

in linear progression. A musical interval is the distance between any two successive pitches within the piece, typically ranging from unison to octave (Figure 4). The most basic of these intervals is defined as an octave, which corresponds to a 2:1 ratio. For instance, we perceive a pitch at 110 Hz to be an octave below a 220 Hz, both of which represent the note ‘A’ [15]. Human beings perceive these ratios to be the same pitch, only at a higher or lower frequency, allowing for the cyclical nature of the scale. We can therefore identify the interval between any two successive notes based upon this scale. While it is not unheard of to have music that utilizes other pitches not represented on the chromatic scale (this is a practice that is observed in many traditional forms of music in the eastern hemisphere), this scale is the backbone of Western music.



Figure 4 - Visual representation of musical intervals ranging from unison to octave

The first reason for this selection comes from the realm of aural skills, in which it is common to use musical intervals as a way to identify differences between eras [16]. Though there are a number of features which are often cited when it comes to aurally distinguishing between eras, intervals are almost always presented as evidence in such efforts, and their status as a cornerstone of music theory make them an obvious answer to our query. Using the musical intervals as features in isolation also provides us with the ability to determine how well it alone can be used to distinguish the era. Secondly, we found that the basis of intervals is an excellent building block upon which to build a generative system, which will be touched upon in greater detail later in our discussion.

3.1.3 Pre-Processing

Once all of these important determinations had been made, it was time to clean the data, and extract the features that had been decided upon. The first step was to collect the data to be used. Though the available pool of **kern scores are not as vast as desired, we were able to accumulate 262 unique pieces of classical music from a variety of eras (Figure 5)

through two Humdrum databases. It is worth noting that the distribution of data entries between these eras were not even across all classes, as there are far less pieces of pre-baroque music that have been encoded using ****kern** format than that of eras such as the classical or romantic era, which feature much more notable composers and pieces which have endured the test of time.

Class	Number of Data Entries
Medieval	10
Renaissance	26
Baroque	77
Classical	50
Romantic	70
Modern	29
Total	262

*Figure 5 – Distribution of ****kern** data between the six classes used within our classifier*

The next step was to extract the features that we desired to use in the classification process. This was perhaps the most tedious task, though we were able to do so in a Linux command line window with a combination of both the Humdrum toolkit, designed for the ****kern** file format (and other formats following the Humdrum Syntax), as well as Linux pattern matching. In the end, we stored the number of times each individual interval appeared and set it as a ratio against the total number of musical intervals encountered.

We appended these ratios (Figure2), along with the era with which the piece is categorized (Figure 1), to the end of an **.arff** (Attribute-Related File Format) file with appropriate headings. Doing this in a loop, we were able to create one file with all 262 musical scores represented. It is with this document that we begin our classification.

3.2 Classification

Classification is an umbrella term to define the task of separating data into distinct categories, and as such there are a large variety of methods that can be implemented in order to achieve the same goal. It became obvious that we would need to test our dataset

with a variety of these classification methods in order to receive the best results possible, and we began work on feeding the data we compiled into five different classification approaches of varying complexity levels.

The two high-level algorithms we utilized in our tests were Multilayer Perceptron (MLP) and Logistic Regression. Based upon an artificial neural network, MLPs use layers of input nodes, output nodes, and two or more layers of hidden nodes to find the most likely path from our input data (comprised of the aforementioned musical interval attributes) to an output identifying whether the data falls within a given class (musical era) or not [13] (Tan). Logistic Regression on the other hand implements a statistical model built upon the probability that a certain piece of data falls within a given class or not. While both of these methods are dichotomous (only have one of two outcomes), they can be used to classify sets with more than two classes when given the dichotomous options of “within the given class” or “not within the given class”.

While Naïve Bayes does not use as sophisticated an algorithm as the above outlined MLP and Logistic Regression models, it is a very well-respected model in the data mining community, and it indeed performs just as well or better than sophisticated models in some instances. The premise of this model is simple, based upon Bayes theorem, which provides a way of calculating the posterior probability of an attribute fitting a defined class [17]. The success of this algorithm lies in the fact that each given attribute is considered independent of one another. As a result, the most probable class is calculated based upon each attribute identified separately, and these probabilities are then multiplied against each other to determine the probability that the piece of data, in this case a musical piece, falls into a given class.

The last two classifiers we utilized, and the simplest of them, into the category of rule-based and decision tree induction predictors. We selected one of each such classifiers, JRip (Rule-Based) and J48 (Decision Tree Induction). JRip uses simple if...then rule structures to split the data into the given classes [13]. J48 uses a similar system within a decision tree structure, where there is a leaf node associated with each of the pre-

determined classes, and classification rules are derived and placed within the ascending nodes as the data is analyzed [17].

3.3 Results

	Medieval	Renaissance	Baroque	Classical	Romantic	Modern	Average
MLP	0.964	0.958	0.854	0.988	0.836	0.996	0.933
LR	0.981	0.951	0.808	0.921	0.885	0.927	0.885
Naïve	0.938	0.931	0.73	0.889	0.853	0.871	0.838
JRip	0.705	0.841	0.73	0.874	0.704	0.836	0.773
J48	0.798	0.777	0.681	0.804	0.741	0.753	0.753

Figure 6: Results of classifiers on our .arff file, based on AUC of ROC graph.

The chart outlined in Figure 6 show a complete picture of the results received from each of the five aforementioned methods of classification. Using an n-fold cross validation approach, the data was partitioned to complete ten iterations of testing. During each iteration of testing, 9/10^{ths} of the data was assigned to act as a training set, used to educate the classifier and build its predictive ability. The other 1/10th of the data was designated to be the test set, used to analyze how well the classifier is able to predict the class the data belongs to. By the end of our ten iterations, all the data has been used as part of a test set and we have a full picture of how accurately the process was able to blindly classifier our data.

In analyzing the results, we chose to focus on the value of the AUC (area under the curve) of a Receiver Operating Characteristic graph as an indication of the success of our classifiers. The reason for this decision is due to the inconsistent number of data pieces between each class represented (Figure 5). The Receiver Operating Characteristic (ROC) Curve maps the True Positive Rate (true positives / all positives) against the False Positive Rate (false positives / all negatives). This produces a curve that will represent how often a piece is mistakenly identified as other than its proper class, rather than produce a true precision rate, which may be skewed as a result of the uneven distribution of data. A perfectly classified set of data would have an AUC of 1.

As seen in the charts, our five classifier models performed at varying levels of accuracy. The most complex algorithm used, the Multilayer Perceptron model, produced AUC rates of .933, while our rule-based and decision tree classifiers lagged behind with AUC rates of .773 and .753 respectively. Perhaps the biggest surprise among our classifiers was the Naïve Bayes model, with an excellent AUC rate of .838, despite the algorithm being quite simple and intuitive.

4. Generation

After analyzing the results of the classifiers, the first step was to determine which classifier was most compatible with our desire to create an algorithmic composition software. On top of providing class predictions, each classifier supplied a model, intended to inform the reader on how it's decision rules were devised. These models are important, as they are the building block upon which we intend to build our music generator. Of the five classifiers, the first two eliminated were the rule-based and decision tree models, JRip and J48. While the classifiers provided positive features, such as easy to understand outputs that outlined the rules used explicitly, it was clear that these approaches were simply not of the same accuracy as their more complex counterparts.

Of our three remaining classifiers, we chose next to eliminate the complex classifiers, Multilayer Perceptron and Logistic Regression. Despite these algorithms statistically doing a better job of classifying the musical scores, the complex models of MLPs and Logistic Regression, based upon mathematical algorithms instead of patterns and rules, did not give a satisfactorily digestible answer as to why the classes were separated the way they were. For this reason, it was difficult to conceive of a way to use these classifiers to inform the generative process of any algorithmic composition software.

We decided to use the knowledge gained from the Naïve Bayes model because it supplied us with a nice middle ground between the previously mentioned choices. It provides an easy, statistical model for us to easily adapt to the generative process. On top

of this, the Bayes model yielded a more respectable AUC value (.838) than the other simple algorithms of J48 (.753) and JRip (.773).

4.1 Method

In perhaps our most contributory work, we move to the generation process of the experiment. The task laid ahead of us was to find a way to utilize the knowledge gained from our Naïve Bayes classifier to inspire the algorithmic composition of music. After consideration of the classifier results and output, we decided to turn our attention to an avenue of algorithmic composition that has been less explored than some others such as artificial neural networks and formal grammars: Cellular automata.

4.1.1 Cellular Automata

The concept of cellular automata (Singular: Automaton) was first proposed by John von Neumann in the 1950's and reached a peak in popularity during the 70's due to John Conway's now famous "Game of Life" 3-D cellular automata model [17]. Based upon the biological cellular replication process, a cellular automata model is represented by a grid of cells, each of which is represented as one of a finite number of states (i.e. "ON" or "OFF"). This grid can be of any finite number of dimensions. The grid progresses in temporally-linear fashion, with each cell shifting states at any given step in time. This shift of the cell states is based upon two factors: the states of the surrounding cells in a pre-determined area defined as it's neighborhood, and a set of transitionary rules which dictate the outcome based on that neighborhood [17]. One of the most famous example of a cellular automata, the Wolfram Elementary Algorithms (Figure 7), adds a new line of cells below the previous generated line with each sequential step in time, with the states of these new cells based upon a neighborhood of the three cells directly above it, and a selected transitionary rule set [18]. With 256 possible rule sets, there are countless possibilities of how the algorithm can compose the sequence of cells, and many produce interesting patterns, such as fractals.

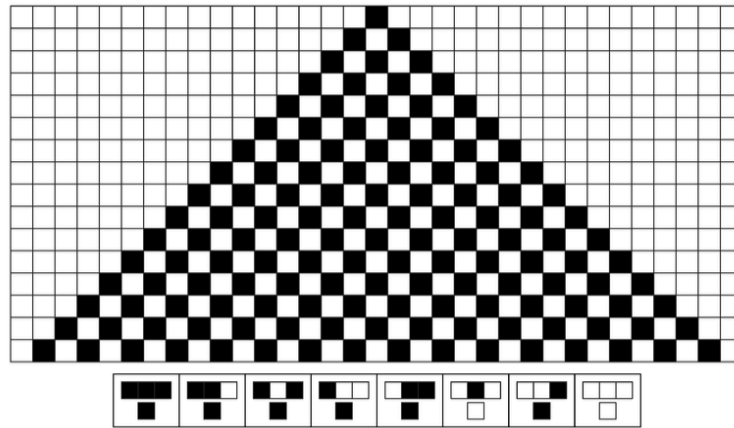


Figure 7 - Rule 250 in the Wolfram Elementary Algorithm Suite,

a popular venture into cellular automata modeling

Rule model's such as Wolfram's provide a unique avenue of exploration for musical composition. The patterns found within these automata rules provide a built-in approach to chaotic music composition. However, those preliminary cellular automata models were only able to create music in an "uncontrolled" way and resulted in music that was not necessarily homogenous with any preconceived style [6]. The next natural step was to create transitional rules that were informed by the true tendencies of music, so as to control the music being generated.

4.2.2 Adapted Musical Model

In an attempt to explore this avenue of musically informed cellular automata, we devised a system inspired by the aforementioned Wolfram Algorithm. Using cells that have one of two states – "On" and "Off" – we are able to interpret a string of these cells as a binary sequence. We chose to map these cells as four-byte binary sequences (16 possible combinations) to the 12 notes of the chromatic circle, with the note C doubled to ease generation given the cyclical nature of the scale. While this system does not currently take into account rhythm, a rest musical character was also encoded for potential future works, as well as terminate and start. A comprehensive look at this binary-mapping is outlined in figure 8.

Binary Representation of Notes

0000	START
0001	C
0010	C#/D ^b
0011	D
0100	D#/E ^b
0101	E
0110	F
0111	F#/G ^b
1000	G
1001	G#/A ^b
1010	A
1011	A#/B ^b
1100	D
1101	C
1110	Rest
1111	TERMINATE

Figure 8 – A table mapping the values of a four-bit binary sequence to the values within the chromatic circle for use in conjunction with cellular automata musical composition

After the groundwork of our cellular automata model was laid out, it was time to create transitional rules inspired by the intelligence gained through our classification process. At the beginning of each transition, a random decimal value between 0.0 and 1.0 was generated. The Naïve Bayes classifier provided a statistical output from which we were able to derive the average probability of any single interval occurring at a given step in time. Figure 9 demonstrates how the probability of a single step interval is represented in this output. We were therefore able to map our randomly generated decimal value to one of the eight interval possibilities. Whichever interval corresponded to the randomly generated decimal value was determined to be the distance between the previous note and our new note. The states of each cell in the four-byte sequence would therefore transition from the previous note's binary representation to a new binary sequence representing our newly found note. In essence, we are generating the interval between the notes, rather than

the note itself. Along with creating more aurally pleasing musical phrases, this helps ease the challenges of representing key signatures within pieces of music.

	Medieval	Renaissance	Baroque	Classical	Romantic	Modern
freqTwos						
mean	0.4837	0.424	0.3562	0.3924	0.2786	0.2111
std. dev.	0.0864	0.1258	0.1272	0.0718	0.0734	0.0964
weight sum	12	28	79	52	72	31

Figure 9 – An example of the statistical output provided by the Naïve Bayes classifier

pertaining to the frequency of stepwise intervals

To help visualize this process, figure 10 provides a mock example of this process. In this example, we are attempting to replicate the medieval era. Thus, the mean frequency values match those discovered by our Naïve Bayes classifier for the medieval era. The decimal value .6197 is randomly generated and mapped within the mean frequencies of the medieval era. It is determined that the decimal value falls within the stepwise interval partition of our chart. Therefore, if we were ascending from the note C, or 0001, we could arrive at D, or 0011.

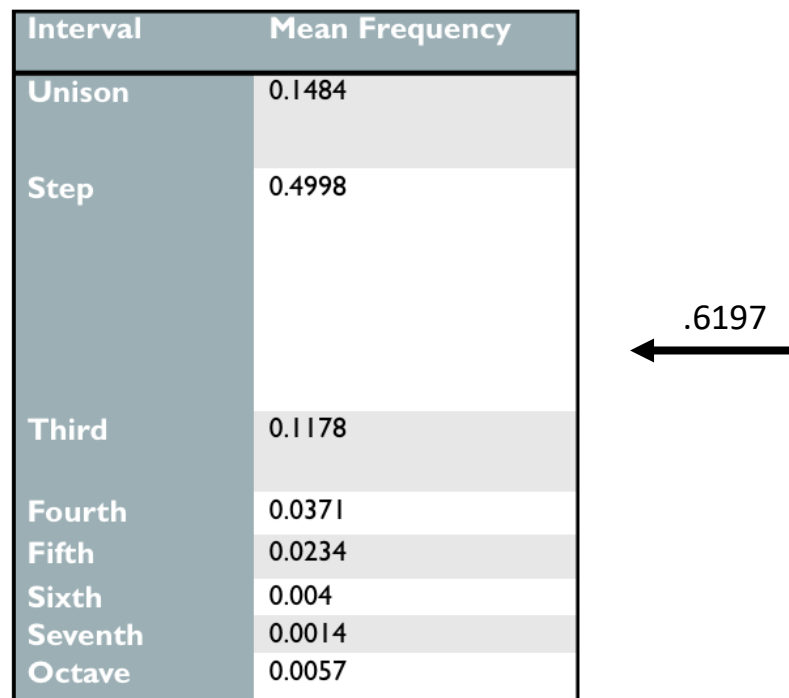


Figure 10 – A visual representation of how a random decimal number is

mapped to the probabilities of each musical interval

To further demonstrate the potentials of this system, the software gives the user the ability to select which era of music they wish to replicate. At the click of a button, the system is able to swap the statistics used in transitionary rule generation to those indicated by the Naïve Bayes output to correspond with the user's indicated era, so as to encourage the system to follow the tendencies of the desired era. This feature helps the software stand out and puts to use the predictive power of our classification approach to rule generation.

The last feature we implemented was a range-check system. In preliminary testing, we found that allowing the note to change in ascending or descending fashion on a 50-50 basis, while relatively common sight within the world of music, was not controlled enough for our experiment, as the true randomness allowed for many algorithmic compositions to get out of hand in terms of range. We therefore found the average distance between the highest note and lowest note within an era of music and dictated that the composition software stays within that range when composing. This allows music that has traditionally had more range to flourish in this sense, while static pieces from earlier eras stick within a more contained range of notes.

4.2 Results Analysis

The result of our efforts is a composition software that is able to imitate any one of six distinct eras of classical music. The system linearly produces a sequence of successive notes based upon the intervals between the previous note and the newly generated note. The pitches are outputted as they are generated using a Java MIDI import at a constant rate that can be changed in the code (currently set to one note every 750 milliseconds).

With the system functioning in the desired fashion, our next step was to analyze just how well our composition software was able to imitate the various classical eras. We chose to implement two different methods of analyzation, to see how well the system was able to reproduce the various eras in both a mathematical and an aural fashion.

4.2.1 Machine Analyzation

In our first of two efforts to analyze the results of our compositions, we used a machine approach closely tied to the ways in which we created the software – classification. While we previously described a ‘n-fold cross verification’ approach during our initial classification process, we decided upon using a ‘test set’ approach for the following exercise. In this approach, we feed the classifier a set of data points known as a training set to develop its knowledge on what distinguishes the different classes, and then feed it a set of data points known as a test set to see how accurately it is able to classify those pieces within the given classes.

To do this, we generated sixty pieces of algorithmically composed music – ten within each era and each piece with a length of 100 notes. We extracted from these compositions the same features we outlined in section 2.1.2, and translated the results into an .arff file mirroring the structure of our previously used .arff file. We then used this file as our test set and provided the file from our initial classification exercise as a training set. We ran these classification techniques on four of the five classifiers used in our original exercise, excluding the Naïve Bayes classifier we used to inform the composition software, as it would provide an unnaturally insightful look into the data, resulting in skewed results. The classifiers’ results are displayed in the chart below (Figure 11).

	Medieval	Renaissance	Baroque	Classical	Romantic	Modern	Average
MLP	0.942	0.9	0.858	0.918	0.754	0.986	0.893
LR	0.978	0.938	0.824	0.946	0.836	0.998	0.92
JRip	0.852	0.753	0.662	0.816	0.582	0.786	0.742
J48	0.812	0.757	0.757	0.8	0.678	0.826	0.772

Figure 11: The results of our algorithmic compositions being classified

*against a training set of the original 262 **kern scores*

The classifiers performed quite well in determining the era which our composition software was attempting to replicate. In fact, the classifiers success rates were nearly identical to the success rates they experienced with traditionally composed pieces of music,

with their short comings being seen in the same categories. The only classifier that saw significant changes in performance was that of the logistic regression approach, which saw the average ROC percentage jump from .885 to .92. These results alone are highly encouraging.

4.2.2 Expert Analyzation

To double down on our analysis, we decided to take a human approach to the matter as well and consulted a number of experts in music. In total, five scholars of music took part in a survey to determine how well they could distinguish the success of our classifier. The exercise was simple: We generated three 15 second clips of music from each era and presented them together in a random order to the experts. We asked at the conclusion of each triplet for the experts to indicate which era they believed the composition software was meant to represent, and their confidence on a scale from 1-5. We also gave the experts an opportunity to explain how they arrived at that answer, and why they gave the confidence level they did.

The results of our direct method of analysis were not as encouraging as the indirect method. Of our experts, only one was able to predict 50% of the eras correctly, and one failed to correctly predict a single era. The confidence levels of our experts hovered between one and three for most questions, with a distinct increase in both confidence and accuracy with the modern era, which four of our five experts correctly predicted.

5. Discussion

It is clear that the results of our expert analysis tell a very different story than the machine analysis. While our classifiers were able to tell which era of music was being replicated with our composition software to a high level of accuracy, experts in music had a much harder time doing so, with a total success rate of 20% when presented the option of all six eras. Compared to true randomness, which would accurately predict the era 16.6% of the time, this is an improvement, albeit slight.

Because of the nature of the process, it comes as no surprise that our two methods of analysis yielded such different results. This is likely because of the limited scope with which we approached the problem, deciding to focus on a very select number of features, even though the differences in musical styles between the eras is defined by many more features, such as rhythm and harmony (A distinction many of our experts pointed out during their survey), as well as the types of instruments being used in the pieces, which is ignored by using a MIDI output.

5.1 Conclusion

From these results, the most evident conclusion is that there is more work to do. The gap between our two methods of analysis show how far we are from creating a musically homogenous algorithmic composition system. Despite this, it is certainly promising that the features we did choose to use in the experiment yielded such high results in our machine evaluation. This shows that, even if the music is not very aurally identifiable yet, trained AI has the ability to distinguish the differences. This result indicates that the project has potential moving forward, and better results may be achieved by integrating more defining features of classical music.

5.2 Applications

For now, it seems the application of this software lays firmly in the category of ‘composition inspiration software’ that encompasses so much of the work that has been done in the field, though it certainly shows signs that it has the potential to be more. The success of our classifiers in determining which era the piece was meant to replicate indicates that there is a lot of potential in the system, when put to use in the correct fashion. The cellular automata system also lends itself to be used with different classifiers, or perhaps even different types of music, as it has been designed to be adapted to any kind of transitional rule set.

5.3 Future Works

At the end of the study, our thoughts on moving forward are much the same as they were when we began. The prospect of hybridizing the various methods of algorithmic music composition with data mining is a vast well of potential which this study has only begun to scratch the surface of. Based on the experts' opinions that our focus on the feature of musical intervals was not enough to encompass all the characteristics of a classical musical era implies that more hybridization must be done with this system to make it more aurally accurate.

There are a number of avenues that could be explored in the pursuit of improving the system in such a manner. This could include varying the instrumentation based on which era it derives from, factoring into the composition rhythm and dynamics, and creating a two-line system that generates harmonious interval sequences. Another feature that could yield positive results would be to adapt the system to employ an n^{th} -order technique, much like the progression of the *Illiac Suite* [7], where we no longer only consider the last note in our generative process. This would allow the music to flow with more natural phrasing and would allow the intervals to take into account where it appears in the musical phrase. Lastly, improvements could be made to the range-check system implemented in this study, which would go hand-in-hand with the phrasing achieved in the n^{th} -order additions.