SOMphony: Visualization and Comparison of Symphonies Through Application of Time Series on 3D SOM

A Thesis Proposal
Presented to
the Faculty of the College of Computer Studies
De La Salle University Manila

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science in Computer Science

by

CRUZ, Edwardo
DIONISIO, Jefferson
FUKUOKA, Kenji
PORTALES, Naomi

Fritz Kevin FLORES
Adviser

August 14, 2017

Abstract

Symphonies are musical compositions for orchestras usually consisting of several large sections called movements. They are usually composed of three to five movements, depending on the time period and are constructed by many different composers (Libin, 2014). There are five major musical periods namely the Baroque period, Classical Period, 19th Century, Romantic Period, and the 20th Century. By visualizing and determining the relationship of one composition to another, this research will be able to show the visual representation of the styles of the different composers. Similarly, the research will also help identify the influence of composers on one another from one musical period to the next. The study of Azcarraga & Flores (2016) tried to understand their relationship using machine learning, which uses self-organizing maps (SOM) and K-Means to determine clusters, and used the frequency counts in order to determine the comparison, which resulted into a visually comparable image of trajectories. Using frequency count however does not take into account the sequence as well as transitions of music from one after the other. This research aims to use the concept of time series on the clustering of the self-organizing maps. By applying time series on the maps, more accurate results can be made.

Keywords: Machine Learning, Music, Time Series, Self-Organizing Map, K-Means Clustering.

Contents

Chapter 1

Research Description

1.1 Overview of the Current State of Technology

Music has been a part of our culture for hundreds of years, classical music being one of the oldest genre of music. Classical music is rooted in the traditions of early western music and to this day, many people refer to classical music as serious music. Musicians, however, use classical music to refer to music composed during 1750 to 1825, otherwise known as the Classical Era (Bernstein, 1959). The central norms of classical music became established between 1550 and 1900, which is known as the common-practice period. The common-practice period contains a bulk of what we now know as classical music. Under this period there are 3 musical eras: Baroque, Classical and Romantic. Music from the Baroque period are decorated and elaborate, with little to no expression. Works from the Classical era contain repetitive dynamics and clean transitions. In contrast to music from the Baroque period, music from the Romantic period are expressive and emotive, having the ability to paint a vivid picture in the minds of the listeners (Grout, Palisca, 1996); however, Dahlhaus (1981) points out that another musical era existed between the Classical and the Romantic period and he refers to this as the 19th century era. This era serves as the transition period for classical and the romantic period, thus having similarities in style with both eras. After the common-practice period comes the 20th century era, which explores modernism, impressionism, neoclassicism and experimental music.

It was in the common-practice era when symphonies began to be composed. Libin (2014) describes symphonies as lengthy forms of musical compositions which are almost always written for orchestras and are consisting of several large movements. With a history of almost 300 years, symphonies today are viewed as

the very pinnacle of classical music where Beethoven, Brahms, Mozart and other renowned composers were able to find a venue for transcending their creativities and overall influencing them heavily on their music. During the course of the 18th century, the tradition was to write four-movement symphonies (Hepokoski & Darcy, 2006).

Throughout time, different styles have developed, each having features unique to themselves. Tilden (2013) notes the historical influence of composers with each other and how similar the methods of composing classical music are with pop music. Due to these facts presented, symphonies written in the early 20th century may be influenced by the great composers and compositions of the previous eras. Analyzing these musical relationships and comparing one to another is a research area that could be done through both manual and machine learning methods.

McFee, Barringtong & Lanckriet (2012) compare the usage of context-based manual semantic annotation versus their proposed optimized content-based similarity learning framework. With machine learning, the usage of high-quality training data without active user participation and the analysis of more data is possible than with feedback or survey data from active user participation. Human error in the analysis process can also be minimized with machine learning since human-supervised training is minimal. Corra & Rodrigues (2016) shows the analysis of music features using machine learning techniques. According to the MIR community (Silla & Freitas, 2009), the two main representation of music feature content are either audio-recorded or symbolic-based. The former employs the explicit recording of audio files while the latter uses symbolic data files such as MIDI or KERN.

SOMphony, a research paper by Azcarraga & Flores (2016) aims to understand the relationship of compositions between the same composer to denote style as well as to determine if there are similarities between compositions of different periods of music to denote influence between time periods. The research showed the relationships and influences between composers from 5 major musical periods, namely the Baroque Period, Classical Period, 19th Century music, Romantic Period and the 20th century. The research focuses on self-organizing maps (SOM) that are trained using 1-second music segments extracted from the 45 different symphonies. The trained SOM is then further processed by doing a k-means clustering of the node vectors, allowing quantitative comparison music trajectories between symphonies. Their research showed that using self-organizing maps are indeed helpful in visualizing the musical features of a symphony, making it easier to create insights about the relationships within the different pieces and composers. The research concludes that a larger dataset would be needed to confirm whether the approach is indeed valid.

However, SOMphony does not take into consideration the notion of time. In time series data, each instance represents a different time step and the attributes give values associated with that time (Witten & Frank, 2005). To be able to generate time sensitive musical analysis, time series is to be added to the SOM and a new visualization in 3D space would need to be created.

1.2 Research Objectives

1.2.1 General Objective

To incorporate the use of time series in the visualization of symphonies for comparison of similarities

1.2.2 Specific Objectives

1. To include more musical pieces to the data set;
2. To perform feature selection to deermine optimal features to be used;
3. To add in the time series variable;
4. To create a 3D visualization model for the data;
5. To have participants listen and annotate the musical pieces for qualitative data;
6. To verify the results of the 3D SOMphony through the results obtained from the human participants;

1.3 Scope and Limitations of the Research

To expand the data set of SOMphony, the proponents will add an additional 2 symphonies to the existing 3 symphonies for each composer. This will result to a total of 125 symphonies all in all. By having an equal number of symphonies per composer, this maintains a balanced data set for all composers. The criteria for choosing the symphonies to be added would be random to have a better grasp on the general style of the composer.

To be able to generate a self-organizing map, the proponents will use jAudio to extract 600 audio features from musical segments generated from the symphony. For the labelling phase, the proponents will classify the selected features by composition because the research focuses on comparing different compositions and comparing eras or composers. The 600 audio features would be trimmed down through feature selection. Decision trees would be used to determine the features to be kept since those nodes that are nearest to the root node of the tree would be dubbed as the more important features compared to the others. The proponents have decided to have 20 as an arbitrary value for the features to be used. By selecting only 20 features, the data set would have a uniform number of features for all symphonies and it would also enhance the efficiency of training the SOMs as it would not take a long time to extract 20 features compared to extracting 600.

To create a 3D model to represent the symphony, the proponents will assign each generated SOM to a point in time and will be used to create a graph representing each map in a time series. As a result of using time series, our research will be able to better differentiate symphonies that use similar themes but at different periods of time in the composition.

In gathering qualitative data from human participants, the proponents limit themselves to 50 participants. In the case that the target amount is not reached within two months, the researchers will proceed to analyze the results they have obtained. The participant profile would be people that have experience or familiarity with classical music. The participants would be presented with a 3D graph and two music players. They are tasked to annotate specific regions of the symphony if they are indeed similar. However we do not limit the participants to the specified regions, the participants are free to annotate parts that they believe sound similar.

Similar to SOMphony, the proponents will focus on representation of symphonies using SOMs for the purpose of comparison to other symphonies. Through the data obtained from the human participants, the proponents will be able to validate if the 3D visualization method is enough to represent the entirety of the symphony for comparison.

1.4 Significance of the Research

As Tilden (2013) states, the structure of classical and modern music are very similar, having the verse-chorus structure and modern pop songs are first composed

instrumentally as similar with how classical music is composed. Modern music just takes classical music further by adding in voice and combining the different techniques employed by classical music. As this study focuses on comparing different symphonies and analyzing to see how similar they are, the results of this study will show us trends among composers in terms of their influence on one another in a musical era, the influence one composer had over other composers from a later era, and what the particular style of a composer would look like in the SOM. This research will show whether composers from back then had a lasting influence on music 100 or so years from a particular composers time period. This research can also show if a particular composer has a definite coherent style that is present in his musical pieces by comparing his works.

Some possible future application of the results of this study would include the improvement of existing music information retrieval (MIR) techniques used by music databases. Corra, D. C., & Rodrigues, F. A. (2016)s research shows a possible improvement on automatic music genre classification using symbolic-based music features. Similarly, this research can also be used to further improve the algorithms used by playlist managers for the retrieval of similar songs from music databases using the comparison of the trained SOMs.

The application of time series in machine learning would benefit studies outside of music that incorporates the use of time sensitive data. It can be used in future works regarding traffic modelling, weather monitoring, prediction, and other time sensitive fields.

Chapter 2

Review of Related Literature

This chapter discusses the features, capabilities, and limitations of existing research, algorithms, or software that are related or are similar to the thesis.

2.1 Musical Data Representation and Interpretation

Musical Data Representation and Interpretation			
Authors &Year	Title	Research Problem	Approach
Correa, D. C., & Rodrigues, F. A. (2016)	A survey on symbolic data-based music genre classification	Expanding music database needs more accurate tools for music information retrieval	Symbolic-based music feature are used to train system for genre classification.
Dubnov, et. al. (2003)	Using Machine-Learning Methods for Musical Style Modeling	Predicting and determining musical context based on relevant past sample is very difficult because the length of the musical context varies widely	Two approaches, incremental parsing (IP) and the prefix suffix trees (PST), are used in designing predictors that can handle data with very large length.
Cambouropoulos, E. & Widmer, G. (2000)	Automated Motivic Analysis via Melodic Clustering	Finding similarity in music patterns.	Their method uses differences in pitch-intervals and rhythm as basis for splitting one musical motive (small bits of music) from another.

According to Corra, D. C., & Rodrigues, F. A. (2016), as music grows continuously over time, a constant need for an upgrade to satisfy the number and size of music databases causes the development of more accurate tools for music information retrieval (MIR) . MIR is the research field responsible for the development of algorithms or other computational means for the retrieval of useful information from music and the classification of music based on their categories. The ever increasing research on machine learning, the ever expanding abundance of digital audio formats, the growing quality and availability of online symbolic music data, and availability of tools for extracting musical properties motivate this study on

machine learning and MIR. One of the main problems in MIR involves the classification of music based on their genre which this study tackles. The automatic genre classification of music plays a key role in online music databases where websites or device music engines manage and label music content for retrieval.

Symbolic-based data are music features extracted from symbolic data formats such as MIDI and KERN. In the MIR community, two main representations of music content for MIR research are followed, either the audio-recorded or the symbolic content. Audio-recorded content produce low-level and middle-level features, whereas symbolic content produce high-level features. When analyzing music content, it is preferable to extract more features with the high-level feature of the symbolic content since it is closer to the human perception of music. Due to these reasons, symbolic-based content is used for the research. This research further provides overviews of important approaches regarding music genre classification with the use of symbolic-based music features. The research, as a result, reveals that pitch and rhythm are the best musical aspects to be explored in symbol-based music feature classification that lead to accurate results. Some limitations for further improvement on future works however are present such as the small amount of music dataset used in the research, the bias of using western culture music, and the lack of comparison means for the result of the research due to the lack of previous research works regarding symbolic-based music genre classification.

Dubnov, et. al. (2003) formulated that by using statistical and information theoretic tools, one can capture some of the more fundamental trends in musical scores for further analysis. By applying machine learning on these statistical data, one can derive mathematical models for inferring and predicting to a certain extent of generating a seemingly new work based on the classical pieces of some popular composers.

Their main source of data for extracting musical surface, a collection of notes for the musical piece, comes in the form of MIDI files or musical instruments digital interface. Machine learnings primary purpose in this study is to help gather the appropriate data to perform statistical analysis for the usage of applications such as style characterization tools for the musicologist, generation of stylistic metadata for intelligent retrieval in musical databases, music generation for web and game applications, machine improvisation with or without interaction with humans, and computer-assisted composition.

Predicting and determining musical context based on relevant past sample is very difficult because the length of the musical context varies widely. Large contexts make it very difficult to estimate because the number of parameters, computational costs, and data requirements for reliable estimation increases exponentially. To address this problem, the usage of predictors that can handle

data with very large length is necessary. Two approaches are used to design such a predictor, namely the incremental parsing (IP) and the prefix suffix trees (PST).

The IP algorithm is a lossless coding scheme implying that the application of this algorithm doesnt result to loss of some spectrum of music while the PST is a lossy compression. The IP also makes sure that every transition is included in the parsing of the music while in PST, the method is very selective in that some rare events and events that do not improve transition are not included. IP uses online estimation through instantaneous coding meaning that IP continually searches online on possible ways or methods to estimate or predict the next possible sequence while in PST, analysis is done by batches through file compression. By analyzing the statistical data provided by either algorithm, predictions or estimates for the next sequence can be made.

Cambouropoulos, E. & Widmer, G. (2000) stated that music could be categorized into small bits called "motives" . These motives are extracted from a musical piece by determining which clusters of musical data can be grouped together while maintaining melodic and rhythmic coherence. This is achieved by representing a melodic segment as a series of notes while minding musical closeness.

Their paper outlines a method that uses differences in pitch-intervals and rhythm as basis for splitting one musical motive from another. For example, two segments can be considered similar if they share a certain number of component notes or intervals using approximate pattern matching. The segments can also be considered similar if they contain shared elements at different pitches. However, this would require a more advanced pattern matching and data structure.

2.2 Music Visualization

Musical Visualization			
Authors & Year	Title	Research Problem	Approach
Azcarraga & Flores (2016)	SOMphony: Visualizing Symphonies Using Self-Organizing Maps	How influential are composers and their symphonies back in the early musical eras?	Construct 2D SOM trained using k-means clustering to construct visual maps for comparison of symphonies.
Azcarraga, A., Caronongan, A., Setiono, R., & Manalili, S. (2016)	Validating the Stable Clustering of Songs in a Structured 3D SOM	Will constructing the classic 2D SOM as a 3D map be feasible, with the learning algorithm still the same as the 2D map?	The 3D map is designed as a $3X3X3$ cube with $9X9X9$ nodes. The cube is divided into one core cube and 8 corner cubes. The Euclidean distance from core to each corner represents the quality of the different categories or genres.
Foote (1997)	Visualizing music and audio using self-similarity	Is it possible to display the acoustic similarity between any two instants of an audio file as a two-dimensional representation	Audio similarity is computed by parameterizing them into MFCCs and getting the autocorrelation of two MFCC feature vectors V_i and V_j that were derived from audio windows.

Azcarraga & Flores (2016) made a paper about Visualizing Symphonies using Self Organizing Maps in order to know whether the music of the a certain composer and certain century is influenced by their past counterparts. In the map, there are different parts of it that represents a unique sound. Every time a specific pitch is hit by the music, a line is drawn until the music of a certain composer is finished. The study used traditional machine learning algorithm in order to know whether there is similarity across each century composers. Basically, the study counts the frequency of a certain pitch sound and summarizes it in order to compare it with other composers.

In order to make a deeper analysis of the study. A new algorithm and variable is going to be used. In the study of this thesis paper, the researchers will add a new dimension and the new variable is a time. The time where a specific pitch will now be important in comparing it to the other pitches of the composers. With this, Time Lapse Algorithm is going to be used in order to summarize data with the time included.

Azcarraga, A., Caronongan, A., Setiono, R., & Manalili, S. (2016) presents a variant of the classical 2D SOM that is stable with the general clusters not moving around on every training phase. A structured 3D SOM is an extension of a 2D Self-Organizing Map to 3D with a predefined structure. In their research, the 3D map is represented as a 3x3x3 cube with 27 sub-cubes of the same size. Each sub-cube is further divided into 9x9x9 nodes. The structured 3D SOM is a collection of one distinct core cube in the center and 26 exterior cubes surrounding it, hence summing to a total of 27 sub-cubes. Alongside 3D SOMs built in structure, the learning algorithm used in this 3D SOM includes a four-phase learning and labelling phase. The first phase of training involves the semi-supervised training of the core cube. The second phase involves yet another semi-supervised training, but for the eight corner cubes. The third phase involves training the core cube again, but the training will be unsupervised. The fourth and final phase will be the labelling phase. This phase involves the uploading of the music files into the cube and labelling them accordingly. The music dataset used in this research includes songs from 9 genres: blues, country, hip-hop, disco, jazz, metal, pop, regae, and rock. Each genre has 100 songs, thus summing to a total of 900 songs.

SOM is usually represented as a 2D map with the input elements being similar to the input environment. This research verifies that designing the SOM as a 3D map is very feasible, with the learning algorithm still the same as with the 2D map. By extending the SOM from 2D map to 3D, the map is further distinguished into the sub-cubes: eight corner cubes and one core cube in the center. Each corner cube represents a music genre while the core cube represents the song itself. The 3D SOM will be able to identify the quality of the different categories or genres of music albums based on a measure of distortion values of music files with respect

to their respective music genres. Distortion value is measured by the Euclidean distance between the core cube and a corner cube.

Foote (1997) presented a paper on Visualizing Music and Audio using Self-Similarity. In this paper, the acoustic similarity between any two instants of an audio file is calculated and displayed as a two-dimensional representation. Structure and repetition is a general feature of nearly all music, with parts resembling certain parts of the song that came before it. This paper presents a method of visualizing the structure of the music by its acoustic similarity or dissimilarity in specific instances of time through grayscale gradation patterns.

Before getting the similarity measures, the two instants are first parameterized into Mel-frequency cepstral coefficients (MFCCs) plus an energy term. The similarity measure $S(i, j)$ is computed by getting the autocorrelation of two MFCC feature vectors V_i and V_j that were derived from audio windows. A simple metric of vector similarity S is the scalar product of the vectors. A better similarity measure can be obtained by computing the vector correlation over a window w. This captures the time dependence ofthe vectors. To have high similarity measure, the vectors must not only be similar, but their sequence must be similar as well.

Given the similarity measures $S(i, j)$ computed for all window combinations, an image is constructed so that each pixel at location (i, j) is given a grayscale value proportional to the measure. The maximum similarity measure is given maximum brightness. Visually, regions of silence or long sustained notes appear as bright squared on the diagonal. Repeated figures such as choruses and phrases will appear as bright off-diagonal rectangles. If the music has a high degree of repetition, it will show up as diagonal stripes or checkerboards that are offset from the main diagonal. Longer audio files would result to larger images due to the rapid rate of feature vectors. To reduce the image size, the similarity is only calculated for certain time indexes and since S is already calculated at window size w, the paper only looks at time indexes that are an integer multiple of w.

Chapter 3

Research Methodology

This chapter contains phases and activities that will be performed to accomplish the research. The phases listed here will be arranged sequentially unless otherwise stated.

3.1 Research Activities

3.1.1 Concept Formulation and Review of Related Literature

This phase will concern the consolidation of the thesis requirements such as the objective of the research, the research problem to be tackled, and the scopes and limitations of such research. Literatures related to 2D and 3D self-organizing maps, music feature classifications, k-means clustering, and machine learning will be part of the Review of Related Literature.

3.1.2 Data Gathering

This phase will concern the gathering of the additional symphonies to be used for the research to provide more reliable results. Aside from the music dataset used in SOMphony, 2 symphonies will be added to each composer, summing up to a total of 5 symphonies per composer and 125 in all. The proponents have decided to maintain 5 symphonies per composer so that the data set will be balanced.

The process of selecting which symphonies to be added would be by random to have a better grasp of the general style of the composer. The audio files would be retrieved from online sources. The researchers would not take into consideration the file type and bitrate of the audio files since music data that is free for use is limited.

3.1.3 Pre-processing

To start pre-processing, the audio files would be converted into wav files in preparation for splitting. WaveSplitter will be used in splitting the audio file into 1 second segments at intervals of 0.5 second. These segments would undergo feature extraction using jAudio. The result would be an xml file containing all the features determined for each segment. The researchers would then run RegEx to extract the unnecessary text in preparation for labeling. Since the proponents would have supervised learning, the data needs to be labelled according to their composer, composition and file name.

3.1.4 Feature Selection

In this phase, the proponents will trim down the 600 features that jAudio has extracted. Using decision trees, the top 20 nodes will be selected as the top 20 features. The proponents have decided to have 20 as an arbitrary value for the features to be used. By doing feature selection, the data set would have a uniform number of features for all symphonies and it would also enhance the efficiency of training the SOM. The tree model produced after feature selection is what would be used in training the SOM.

3.1.5 Visualization

In this phase, the proponents will trim down the 600 features that jAudio has extracted. Using decision trees, the top 20 nodes will be selected as the top 20 features. The proponents have decided to have 20 as an arbitrary value for the features to be used. By doing feature selection, the data set would have a uniform number of features for all symphonies and it would also enhance the efficiency of training the SOM. The tree model produced after feature selection is what would be used in training the SOM.

3.1.6 Performance Evaluation and Human Evaluation

In this phase, the proponents limit themselves to 50 participants. The participant profile would be people that have experience with classical music. In the case that the target amount is not reached within two months, the researchers will proceed to analyze the results they have. The participants would be presented with a 3D graph and two music players. They are tasked to annotate specific regions of the symphony and verify if they are indeed similar. However we do not limit the participants to the specified regions, the participants are free to annotate parts that they believe sound similar.

3.1.7 Documentation

This phase will be done all throughout the whole research timeframe. This will include taking down notes on observations and findings during experiments and during the review of related literature, writing related technical documents, and the research paper itself.

3.1.8 Calendar of Activities

Table 3.1 shows the time table for the activities involved with the research. The numbers represent the number of weeks worth of activity. The symbol represents the number of weeks allotted for the month.

References

Bernstein, L. (1959). *Young people's concert: What is classical music?* Amberson Holdings LLC. Retrieved from https://leonardbernstein.com/lectures/television-scripts/young-peoples-concerts/what-is-classical-music