B4

Visualizing Symphonies Using 3D Self-Organizing Maps

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

Feburary 30, 2014

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, Laurence, 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. Azcarraga & Flores (2016)s study 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 Seires, Self-Organizing Map, K-Maps.

Contents

Chapter 1

Research Description

1.1 Overview of the Current State of Technology

manual techniques for music extraction

Music has been a part of our culture for hundreds of years, classical music being one of the oldest genre of music. McFee, B., Barringtong, L., & Lanckriet, G. R. G. (2012)~~s research~~ compares the usage of manual technique versus their proposed metric 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, D. C., & Rodrigues, F. A. (2016)~~s research~~ shows the analysis of music features using machine learning techniques. According to the MIR community (Silla, C. N., Jr. & Freitas, A. A., 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.

for what purpose?

Throughout time, different styles have developed, each having features unique to themselves. Imogen 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.

symphony

Azcarraga & Flores (2016) conducted a research 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. By choosing three musical pieces from each of the three composers in each period, the research was able to produce quantitative comparisons of the music trajectories between pieces of the same composer and pieces within the same time period or on another. The research extracted the frequency count for each music segment lasting for a second in order to compare and analyze what pitches are frequently used by a specific musician in his piece.

Self organizing map was used, wherein the map was divided into 21 clusters and each cluster represents a group of similar sounds. 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; however, the data set used only contained 45 compositions and therefore may not be representative of the entire collection of symphonies. There are specific cases wherein similar looking maps are generated from two musical pieces that sound nothing alike. These two pieces may have similar features, but since the maps are not time sensitive, they would end up looking quite similar.

maintain

Consistency (dash or with dash?)

was this noted by the authors in the paper?

find in symphony paper

According to Xu et al. (2017), time series is a type of serial data which includes equally divided points in time order. With the data series, timing of the pitch is now considered and is the main basis in evaluating results is now not only based on the frequency of each pitch. To make a more precise analysis and understanding of the results, time series is added to the study to improve that a particular music is similar to the other. With this new concept in analyzing the similarities of the musical compositions of each composer by counting frequency of each pitch, there is no study wherein time series is considered in evaluating the different kinds of music with visualization.

check grammar

sentence construction needs improvement

prove?

1.2 Research Objectives

To ~~incorporate the use of time series in~~ comparing the given ~~set of musical pieces~~

I think there is a better way to write

1.2.1 General Objective

To incorporate the use of time series in comparing the given set of musical pieces which algorithm are you referring to?

1.2.2 Specific Objectives

which previous version

1. To do a performance evaluation on the algorithm versus the previous version without time series;To include more composers and musical pieces to the data set;
2. To determine optimal features to be used;
3. To add in the time series variable to the current existing visualization; symphony
4. To do a performance evaluation on the algorithm versus the previous version without time series;
5. To have participants listen and compare the musical pieces for qualitative comparison of symphonies;
6. To compare the results of the algorithm and the results of the human participants;

repetition

specify our "new" algorithm

1.3 Scope and Limitations of the Research

The research will focus on inferring similarities and trends between different symphonies of multiple composers within different eras by visualizing each piece as a 3 dimensional graph. 3D

The symphonies used in this research will include influential works from 5 musical eras specifically, Baroque Era, Classical Period, 19th Century, Romantic Era and 20th Century with each era being represented by 3 composers with at most 5 pieces per person. Only orchestras or composers that are considered as influential during its era will be included in the data set.

To be able to generate a self-organizing map the research will use jAudio to extract audio features from musical segments generated from the orchestra. In addition, for the purpose of increasing accuracy, this paper will use 600 features, as opposed to 78 suggested by Azcarraga & Flores (2016), to represent each segment in order to accurately plot it on the graph and better differentiate similarly sounding samples. Each music sample will contain 1 second of playback from the composition with each successive segment beginning 0.5 seconds after the last. The resulting data will be processed through k-means clustering to generate a point in the map that best represents the sound clip. As a result, the map is organized into partitions that denote similarly sounding music segments.

3D

To create a 3 dimensional model to represent an orchestra, the research will assign each generated self-organizing map 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, the paper will be able to better differentiate orchestras that use similar themes but at different periods of the composition.

To evaluate the performance of the method proposed in this paper, the paper will generate 3D self-organizing maps using process previously discussed and 2D self organizing ~~maps using~~ the approach used in ~~the paper of~~ Azcarraga & Flores (2016). Using the orchestras used in their research, the paper will review whether musical pieces deemed as similar using the 2D self-organizing map are also similar when visualized ~~in a 3D self~~ organizing map. In order to calculate whether 2 self organizing maps are similar the paper will calculate the euclidean distance between clusters in the map in order to quantify it in percentages.

which paper

our paper/research

provide be… = our

1) the proponents

2) our proposal

In order to gather qualitative data the paper will have people who are members of an orchestra listen to 5 music samples deemed by the 3D self-organizing map to be similar and 5 that are dissimilar. In total, 50 participants will listen to 10 3-minute-long music samples. The resulting data, when compared to the results generated by the 3D SOM, will help give insight into the actual effectiveness of the papers discussed visualization method.

Similarly to the paper of Azcarraga & Flores (2016), this research focus on representation of symphonies using self-organizing maps to generate an accurate depiction of a musical piece for the purpose of analysis or comparison.

we need to state "symphony" in page 2

"Symphony" isn't discussed previously

1.4 Significance of the Research

Objective

By applying machine learning to the study of music by different composers, as with Azcarraga & Floress (2016) paper, the study aims to reveal trends on how one composer influenced anothers work through ~~the~~ visualization of the self-organizing maps in SOMphony but now with an added time series variable to further improve the accuracy of the trained maps.

As Imogen 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. 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 particular composer would look like in a self organizing map. 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.

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.

Ahh ba talaga?

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

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.

fix typesetting on double quotes

Cambouropoulos, E. and 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

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 Vi and Vj 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

* provide more details in this Chapter

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 reviewed as part of the Review of Related Literature.

a

3.1.2 Data Gathering

This phase will concern the gathering of the additional symphonies to be used for the research. Aside from the music dataset used in the research of Azcarraga & Flores (2016), 2 symphonies will be added to each composer, summing up to a total of 75 symphonies. This phase will also include the division of the additional symphonies into 1 sec music segments in preparation for music feature extraction.

data from where you will get your data (data source)

seconds

ask for data where he got the audio files

* Also discuss human participants from UP! (music experts)

place in the algorithm formulation?

3.1.3 Development

In this phase, the concept of time series will be incorporated to the algorithm for training the SOM. The system for the visual representation of the trained 3D SOMs will also be developed. This 3D SOM visualization will look similar to Azcarraga, A., Caronongan, A., Setiono, R., & Manalili, S. (2016)~~s research~~ work but with the altered algorithm for time series.

3.1.4 Training

In this phase, training of the symphonies into 3D SOMs will be done. JAudio will still be used for the music feature extraction and algorithm will for clustering will be modified accordingly based on the results of development. 600 optimal features among all the features will be selected as a result. The newly added, together with the old symphonies should be trained all over again using ~~the new~~ 3D SOM.

maybe we could say the name of our algorithm

3.1.5 Performance Evaluation and Human Evaluation

In this phase, comparison between the results of the previous 2D SOM and the results of the 3D SOM for the symphonies from the old dataset will be made. This, together with the results of the symphonies from the new dataset will make up the quantitative comparison of symphonies. This phase will also include the evaluation of results from at least 50 human participants with knowledge about music who will listen and give comparisons on specified symphonies. This will make up the qualitative comparison of symphony. Both the quantitative and qualitative comparisons will be used in analyzing and concluding the results of the experiment.

3.1.6 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.

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.

3.2 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.

~~draft~~ typeset in LaTeX

Table 3.1 Timetable of Activities

Activities (2017)	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Concept Formulation and Review of Related Literature	⍰ ⍰ ⍰	⍰⍰ ⍰ ⍰	⍰ ⍰ ⍰	⍰ ⍰ ⍰			
Data Gathering				⍰ ⍰	⍰ ⍰ ⍰ ⍰	⍰ ⍰ ⍰	
Development					⍰ ⍰	⍰ ⍰ ⍰ ⍰	⍰ ⍰
Training							
Performance Evaluation and Human Evaluation							
Documentation	⍰ ⍰ ⍰	⍰⍰ ⍰⍰	⍰ ⍰ ⍰	⍰ ⍰ ⍰	⍰ ⍰ ⍰ ⍰	⍰ ⍰ ⍰ ⍰	⍰ ⍰

Activities (2018)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug

Concept Formulation and Review of Related Literature								
Data Gathering								
Development	? ? ?	? ? ? ?						
Training		? ?	? ? ? ?	? ? ?				
Performance Evaluation and Human Evaluation				? ? ?	? ?	? ? ? ?	? ? ? ?	? ? ?
Documentation	? ? ?	? ? ? ?	? ? ? ?	? ? ?	? ?	? ? ? ?	? ? ?	? ? ?

format in LaTeX

References

Azcarraga, A., Caronongan, A., Setiono, R., & Manalili, S. (2016). Validating the Stable Clustering of Songs in a Structured 3D SOM.

Cambouropoulos, E. and Widmer, G. (2000). Automated Motivic Analysis via Melodic Clustering. Journal of New Music Research, 29(4), pp.303-317.

Corrêa, D. C., & Rodrigues, F. A. (2016). A survey on symbolic data-based music genre classification. Expert Systems with Applications, 60, 190-210.

Imogen Tilden. (2013). What Pop Music Owes to the Musical Masters. Retrieved from https://www.theguardian.com/music/2013/jan/24/what-pop-music-owes-classical-masters

Libin, Laurence. (2014). Symphony. Retrieved from https://www.britannica.com/art/symphony-music

Shlomo Dubnov, Gerard Assayag, Olivier Lartillot, and Gill Bejerano. 2003. Using Machine-Learning Methods for Musical Style Modeling. *Computer* 36, 10 (October 2003), 73-80. DOI=http://dx.doi.org/10.1109/MC.2003.1236474

Foote, J. (1999, October). Visualizing music and audio using self-similarity. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)* (pp. 77-80). ACM.

Silla, C. N., Jr. , & Freitas, A. A. (2009). Novel top-down approaches for hierarchical classification and their application to automatic music genre classification. In Proceedings of the IEEE international conference on systems, man and cybernetics, San Antonio, Texas (pp. 3499–3504)
.

McFee, B., Barringtong, L., & Lanckriet, G. R. G. (2012). Learning Content Similarity for Music Recommendation. In IEEE Transactions on Audio, Speech, and Language Processing, 20 (8), 2207-2218.