Application of Deep Learning to Classification of Hindustani Classical Music Scale Patterns

Anshita Saini

Interlake High School

Mentored by Mr. Peterson (petersond@bsd405.org), Interlake High School

Abstract

This proposal explores the potential classification of Hindustani classical music ragas, or musical patterns distinguished by distinct movements of notes. Traditionally, Hindustani music is passed down orally, and ragas may develop and adopt new patterns over thousands of years. The desired outcome of this project is the creation of a classifier that can distinguish between ten different ragas with over 90% overall accuracy. Should a raga change enough to be completely separate from what the classifier has learned about the raga, a new raga classification can be created, resolving conflict between learners of Hindustani classical music between the two similar, yet differently developed ragas. The classifier will be created using a deep learning approach, specifically a two-dimensional convolutional neural network. A two-dimensional convolutional neural network uses various algorithms to recognize prominent features in a set of data, key to classifying different datasets. The underlying methodology behind a neural network is applicable to the classification of ragas in Indian classical music, as each raga is differentiated by certain, specific patterns. Such identification is what the aforementioned architecture is primarily built for. Based upon this approach, an experiment can be carried out involving the creation of a two-dimensional convolutional neural network to accurately classify between the various ragas. The machine learning model will be evaluated utilizing 10-fold cross-validation with a complete confusion matrix combined across all folds.

Investigative Proposal

Recent research has explored the idea of identifying patterns within audio data. Specifically, machine learning and neural networks have been applied to classify musical genres. This paper proposes the creation of a two-dimensional neural network that will be trained and tested to classify ten different ragas, or complex patterns in Hindustani classical music.

Music Classification Background

Many projects have previously attempted to classify musical genres using deep learning, such as identifying popular music versus rhythm and blues music (Dwivedi, 2019). These projects have been largely successful, leading to the question of classifying music from other cultures, such as Hindustani classical music. While Western music has these different genres that could potentially be classified computationally, at a first glance, one of the main aspects of Indian classical music that could be classified in this way is the different *ragas*. Ragas are often compared to their Western equivalent of scales, a concept fundamental to all music since the times of ancient Greece (Sadhana, n.d.). In the same way that Western music includes major, minor, and pentatonic scales, there are ten parent scales, known as *thaats*, in Indian classical music, and these are what *ragas* are derived from. Since Indian classical music was developed in Vedic hymns of ancient Hindu temples two thousand years ago, musicians have been creating many different ragas based upon the ten original thaats (Sadhana, n.d.). Ragas differ slightly from scales in that the ascending and descending scales of a raga may be different, while a Western scale is the same whether ascending or descending. Two ragas can often have the same sets of notes with extremely unique ascending and descending scales. Moreover, while Western scales are generally played with uniform emphasis on each notes, different ragas are differentiated by their individual notes of focus and distinctive sequences or ‘moods’ (Patel, 2017). If a particular sequence from one raga is used in another raga, it may sound completely out of place to a trained musician. These patterns stand out as a unique aspect of these ragas in Hindustani classical music, attracting the attention of researchers to classify these ragas computationally. Certain challenges lie in classifying this type of music, as it requires consideration of both intrinsic properties, such as notes, and extrinsic properties, such as emotions conveyed (Ross, n.d.).

Deep Learning

First and foremost, the concept of deep learning must be explored in overarching terms to gain a greater understanding of its applicability to classification problems within music. The concept of deep learning implies modelling data with a computational system combining non-linear transformations (“Neural Networks,” n.d.). Like their parent field of machine learning, deep learning models rely on inputs of training data and test data. Neural networks utilize the training data by processing these inputs in a structure relying on hidden layers of neurons. All inputs are ‘weighted’ individually and passed into an activation function, which varies depending on set parameters chosen by the user, resulting in an output (“Neural Networks,” n.d.). Essentially, within the first few layers of a deep learning network, the neurons identify simple patterns within the input, and later form more complex patterns within until the network arrives at a final classification. The final classification of the network implemented in this investigation outputs a set of probabilities that a certain input variable correlates to a potential output. Deep learning has been applied to many fields including natural language processing, speech recognition, and audio recognition, which will be the primary pursuit of this investigation (“Neural Networks,” n.d.).

Applicability of Two-Dimensional Convolutional Neural Network

Knowing that chosen ragas will have certain patterns and emphasized notes that distinguish them from one another, it is now imperative to identify how a two-dimensional convolutional neural network could specifically be applied to the fascinating challenge of the classification of Indian ragas. Firstly, the methodology of a two-dimensional network for this project rather than a one-dimensional neural network must be appreciated. In previous projects classifying music, audio files were converted into spectrograms, or representations of the frequencies of the file, rather than one-dimensional frequency-time graphs, as these spectrograms were the most holistic representation of the ragas (Dwivedi, 2019). In most research, one-dimensional convolutional neural networks are often implemented for time-series data (Ackermann, 2018). Thus, two-dimensional convolutional neural networks could be applied for raga classification, as they are suited for the RGB inputs of spectrogram processing projects, while one-dimensional networks would be applicable to time-series data. Moreover, two-dimensional neural networks are especially effective in finding features from short segments of a dataset (Ackermann, 2018). Thus, a two-dimensional convolutional neural network proves to especially useful in classifying ragas, as the patterns of the melody within shorter, fixed-length portions of a raga dataset are more significant for classification, while the location of a particular point in the audio dataset does not matter (Kumar, 2014).

Current Work

Because computational musicology in the field of Hindustani music is a niche field, research has been limited upon this topic. Previous work has utilized a singular Recurrent Neural Network to classify Carnatic ragas (Kumar, 2014). In addition, hybrid architecture with a paralleled convolutional neural network and recurrent neural network has been applied to the classification of Experimental, Folk, Hip-Hop and Instrumental genres (Dwivedi, 2019). Moreover, research has additionally been completed on the use of Bayesian Net and Naive Bayes classifiers; however, all techniques researched only yielded an 83% accuracy after discretization (Patel, 2017). Lastly, n-gram and pitch class profiles of ragas were combined in another classification experiment, still only yielding a 83.89% accuracy rate, and this project too was largely aimed towards Carnatic music (Sharma, 2015). No previous projects have reached the goal of a classifier with 90% accuracy between Hindustani ragas, rather than Carnatic.

Further Discussion of Investigative Approach

The earlier introduction to two-dimensional convolutional neural networks describes their applicability to the classification of images. Using a two-dimensional convolutional neural network will be particularly suitable for analyzing spectrograms, which are like images with distinct patterns (Dwivedi, 2019). Figure 1 displays the architecture of the network.

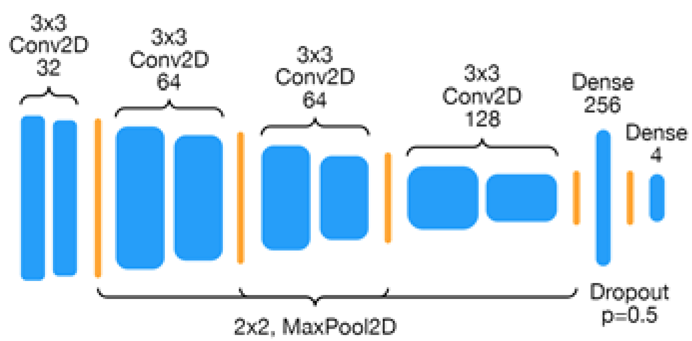

Figure 1. The architecture of the 2D convolutional neural network that will be created.

Having identified that the approach of this investigation will be a neural network trained and tested upon a dataset of spectrograms, the creation of the dataset requires further discussion. Data will be collected through a web application, with a Start/Stop button functionality to save a .wav file and text boxes to record notes on the singer and raga being sung, which will be saved in a .json file format. Because a visually clear spectrogram must be identified from the audio file, any computational library would not function properly if the audio file in question has an excessive amount of ornamentation or a lack of clarity. This issue is especially prominent in the field of Hindustani classical music, as the majority of ragas in Indian music are often sung with accompanying instruments, such as the tabla or a harmonium. These conflicting layers of instrumental versus vocal music within potential datasets could make it extremely difficult to obtain numerical data. Thus, the input data must be limited to data that is solely vocal, without instrumental supplements that could interfere with the main, necessary pitches, or data that is solely instrumental. Lastly, the scope of the investigation was defined before creating the neural network. With an extremely expansive set of variations over its history, Hindustani classical music contains many ragas that have extremely similar patterns and are even difficult for a trained musician's ear to distinguish between, much less a computational structure, no matter how complex it is ("Indian classical music," 2016). Thus, the investigation question was limited to a certain number of ragas – in this case, ten ragas were chosen, derived from ten of the aforementioned parent *thaats* to cover the broadest scope within the comprehensive tree of Hindustani ragas.

Resources

The resources needed for this process are free Python libraries, including Keras, an open-source neural-network library, supporting convolutional networks ("Keras, n.d). In addition, SciPy was used to generate spectrograms (Scipy, n.d.). These libraries will be run on a home laptop, specifically the X1 Carbon Gen 7. The main mentor will be Dan Peterson, computer science teacher at Interlake High School and the secondary mentor will be Parker Ruth, undergraduate student researchers at the Ubiquitous Computing Lab at the University of Washington with extensive experience in signal processing and deep learning. However, creation of the dataset will require recording equipment and experienced Hindustani classical singers. Recording materials will be acquired based upon the budget and will likely require a semi-professional mic, as well as converters depending upon the mic type.

Goals

Four main milestones must be achieved to complete this project:

1. Dataset creation
    a. Recording at least 100 sound files, 10 per raga
    b. The dataset must be recorded with five trained Hindustani classical singers.
2. Data processing
    a. Creation of spectrograms from the sound files
3. Creation and training of the two-dimensional convolutional neural network
4. Evaluating the neural network

The completion criteria of the project will be classification of ten ragas with 90% accuracy on the testing dataset, evaluated through ten-fold cross validation.

Timeline

A timeline with key deadlines can be found below, created for a six-month period from January to June 2020. The implementation process between these milestones will be documented with regular commits to a public GitHub repository. Sound files for the dataset must be uploaded on the dates listed, and updates to the network with additional data processing programs must be committed with appropriate commit labels to the same repository on deadlines.

Deadline	Deliverable Name	Deliverable Description
01/15/2020	Project begins	Create GitHub repository
01/20/2020	Data collection begins	Contact up to 10 local trained Hindustani classical music singers
02/20/2020	Dataset collection completed	Have at least 100 uploaded vocal sound files of trained singers singing chosen ragas (ten files per raga)

03/15/2020	Dataset processing completed	Generate spectrograms for every sound file Ensure that each spectrogram appears clear
04/15/2020	Neural network creation	Code a 15-layer two dimensional convolutional network using Keras, splitting the training/testing data with a 70:30 ratio
04/30/2020	Neural network training completed	Train the model upon 70% of the files from the dataset
05/10/2020	Neural network testing completed	Test the model upon 30% of the files from the dataset
05/20/2020	Evaluation of model	Evaluate the model utilizing ten-fold cross validation The results must be displayed in a confusion matrix combined across all folds
06/15/2020	Report	A 10-page report with a summary and evaluation of the model will be written

Current Progress and Need for Funding

Background research has been completed on the creation of this two-dimensional convolutional neural network and the architecture has been predefined with fifteen layers (see Figure 1). The ten ragas for the project have been chosen. The ragas and their Western scale equivalents can be seen in Figure 2. However, a dataset needs to be created and the neural network must be trained and tested. Funding from the MIT THINK Scholars Program will allow these proposed goals to be achieved by providing financial support for hiring trained Hindustani classical music singers to create an appropriate dataset and to acquire appropriate recording equipment to ensure maximum accuracy.

Potential Issues

One issue that may be encountered during implementation is inaccuracy of the neural network due to two ragas being very similar. Should two ragas have similar patterns, as some do in Hindustani music, they could be identified inaccurately. This risk will be mitigated by choosing ragas based on the ten thaats. Because these are ‘parent’ ragas they will be very different, rather than being under the same branch, and will likely be able to be classified accurately by the network. Another issue that could be encountered during implementation is noise in the sound files, obscuring the signal of the voice. This must be mitigated through professionally recorded files and noise reduction through signal processing to focus upon the voice of the singer. Lastly, a potential risk could be that the network might not be able to accurately classify files from different singers singing the same raga. Different singers have different methods of singing ragas based upon individual interpretations. This will be mitigated through hiring trained Hindustani singers, as learners of Hindustani music in the local area generally learn similar compositions and will not sing in starkly different manners.

Project Budget

The table below displays the costs of this project.

Item	Amount	Cost	Link to supplier
Trained Hindustani singers - will sing for two hours	5 - $60 per singer for one hour	$600	http://localservices.sulekha.com/entertainment/singers-in-bellevue-wa
Blue Yeti USB Mic	1	$100	https://www.amazon.com/Blue-Yeti-USB-Microphone-Midnight/
	Total	$700	

Personal Interest

My personal interest in this project stems from challenges in learning Hindustani classical music. One of the tasks I struggled with in my exam for my Junior Diploma in music was identifying *ragas* by ear. During the time I was struggling with this, I had just started taking AP Computer Science and found that computers could be applied to many challenges that humans face in their daily lives. This led me to consider the question of whether a computer could do what I could do by classifying *ragas* and whether it could perhaps do this task better than I could. Currently, I have research experience through my work on the Exercise-Rx project during my continuing internship at the Ubiquitous Computing Lab at the University of Washington. The research project aims to quantify upper body movement exercises through Doppler sensing. For this project, I created a one-dimensional neural network as a binary classifier between two exercises using time-gyrometer data. In addition, I aided in the creation of a dataset for a two-dimensional neural network that aimed to distinguish between exercises based on spectrograms of Doppler motion, giving me experience in signal processing and spectrogram creation and reminding me of my original idea of classifying ragas that I had thought of two years ago. This experience is further applicable to this project as I now know how to process a signal and effectively focus spectrograms to the correct resolution. However, I have not created a two-dimensional network before, though I do have experience with a one-dimensional network.

In general, I am interested in this project because I believe that deep learning can greatly simplify and aid in human tasks and combining it with music and art is something that not many people consider at a first glance. My general career goal is to use computer science, specifically artificial intelligence, to create products or methodologies that aid people in their day-to-day life, and my research aligns with this long-term goal.

References

Ackermann, N. (2018, September 14). Introduction to 1D Convolutional Neural Networks in Keras for Time Sequences. *Medium*.

Dwivedi, P. (2019, March 27). Using CNNs and RNNs for Music Genre Recognition. Retrieved August 15, 2019, from https://towardsdatascience.com/using-cnns-and-rnns formusicgenre-recognition-2435fb2ed6af.

Indian classical music roots and ragas - Times of India. (2016, September 29). Retrieved August 18, 2019, from https://timesofindia.indiatimes.com/Indian-classical-music-roots-and ragas/articleshow/54540906.cms.

Keras: The Python Deep Learning library. (n.d.). Retrieved August 29, 2019, from https://keras.io/.

Kumar, V., Pandya, H., & Jawahar, C. (2014). Identifying Ragas in Indian Music. *22nd International Conference on Pattern Recognition*.

Neural Networks and Introduction to Deep Learning. (n.d.). *WikiStat*.

Patel, E., & Chauhan, S. (2017). Raga detection in music using supervised machine learning approach. *International Journal of Advanced Technology and Engineering Exploration*, 58–67.

Ross, & Cheri, J. (n.d.). Identifying Raga Similarity Through Embeddings Learned From Compositions' Notation. *IBM Research India*, 515–522.

Sadhana. (n.d.). What is a Raga? Retrieved September 28, 2019, from https://raga hindustani.com/Scales1.html.

SciPy.org. (n.d.). Retrieved September 18, 2019, from https://www.scipy.org/.

Sharma, Hiteshwari, & Bali, R. S. (2015). Comparison ML Classifiers for Raga Recognition. *International Journal of Scientific and Research Publications*, *5*, 1–5.