Spring 2018

Music Similarity Estimation

Anusha Sridharan
San Jose State University

Music Similarity Estimation

A project

Presented to

The Faculty of the Department of Computer Science

San JoséState University

In Partial Fulfilment

Of the Requirements for the Degree

Master of Science

by

Anusha Sridharan

May 2018

The Designated Project Committee Approves the Project Titled

Music Similarity Estimation

by

Anusha Sridharan

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

San José State University

May 2018

Dr. Teng Moh Department of Computer Science

Dr. Chris Pollett Department of Computer Science

Dr. Robert Chun Department of Computer Science

ABSTRACT

Music Similarity Estimation

by Anusha Sridharan

Music is a complicated form of communication, where creators and culture communicate and expose their individuality. After music digitalization took place, recommendation systems and other online services have become indispensable in the field of Music Information Retrieval (MIR). To build these systems and recommend the right choice of song to the user, classification of songs is required. In this paper, we propose an approach for finding similarity between music based on mid-level attributes like pitch, midi value corresponding to pitch, interval, contour and duration and applying text based classification techniques. Our system predicts jazz, metal and ragtime for western music. The experiment to predict the genre of music is conducted based on 450 music files and maximum accuracy achieved is 95.8% across different n-grams. We have also analyzed the Indian classical Carnatic music and are classifying them based on its raga. Our system predicts Sankarabharam, Mohanam and Sindhubhairavi ragas. The experiment to predict the raga of the song is conducted based on 95 music files and the maximum accuracy achieved is 90.3% across different n-grams. Performance evaluation is done by using the accuracy score of scikit-learn.

ACKNOWLEDGEMENTS

I would like to thank my project advisor Dr. Teng Moh, for his continuous support, guidance and encouragement throughout this project. I would also like to thank mycommittee members, Dr. Chris Pollett and Dr. Robert Chun for their time and support.

I would also like to thank the library for easy access to conference paper and books to learn about our project. I would also like to thank Ajay Srinivasamurthy and Alastair Porter from Music Technology Group of Universitat Pompeu Fabra, Barcelona, Spain for helping me to access the full carnatic music audio collection from dunya compmusic.

Last, but not least, I would like to thank my family and friends for the encouragement and moral support without which I would not have completed the project successfully.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

Classification music based on genre has gained its popularity in both domains of Music Information Retrieval (MIR) and machine learning in the past few decades. Many music streaming platforms like Spotify, Pandora and Saavn use automated music recommendation services. A huge amount of music is now accessible to the public and this has given rise to the necessity for developing tools to effectively manage and retrieve music that interests end users [1]. The primary step to manage and track these systems is to classify music. Classification of music is important because majority of the listeners will be eager to listen to specific types of music and classification would be beneficial to recommend and promote desired songs to them [3]. Classification can be based on various parameters like genre, emotion and mood. When a person listens to music, he/she recognizes the feel, culture and emotion of the music. He/she cannot relate to acoustics of the sound wave. Due to its acoustic and cultural complexity, music classification is an intricate process. The project aims at classifying western music based on genre and the maximum accuracy achieved is 95% and Carnatic music based on raga and the maximum accuracy achieved is 90.3%.

Music can be classified based on genre, artist, music instrument, emotion and mood. Genre classification involves categorizing music by genre and naming them with labels like ‘blues, ‘pop’, ‘rock’ and ‘classical’. Music emotion classification is of two types: Valence-Arousal scale and Geneva Emotional Music Scale (GEMS). Valence differentiates positive and negative emoted tone and Arousal differentiates high and low emotion. Valence-Arousal scale involves classifying music into four types namely happy, sad, angry and relaxed. GEMS involve classifying music into 9 categories namely power, joy, calmness, wonder, tenderness, transcendence, nostalgia, tension and sadness [18].

Music Genre Classification is a familiar problem in the field of MIR. The approaches commonly used for genre classification use low-level features like Fast Fourier Transform and Mel Frequency Cepstral Coefficient [2]. Fast Fourier Transform converts input signal into frequency representation and this is used to analyze intensities of various pitches. Mel Frequency Cepstral Coefficient involves calculating Discrete Fourier Transform and taking discrete cosine transform of a sound wave [3]. Low level features are widely used because it is easy to extract them and the results they produce in classification system are promising. However, the low-level features are not understood by human listeners or even experts in music. Due to this, a logical relationship cannot be established between these features and the outcome which could be genre or emotion. Mid-level features also contribute in understanding the long-term features. To overcome the limitation of using low-level features, we use mid-level features like rhythm, pitch and harmony. The usage of mid-level features is widely used in problems like cover song detection and query by example. These features are relatable to music but it is challenging to extract them.

On the other hand, Indian classical Carnatic music is disparate from western music. Carnatic music has its origin in southern India. Classification of Indian classical Carnatic music is performed based on raga, tala and artist. Raga is a sequence of swaras (notes) and is defined by arohana and avarohana. Arohana is ascending scale of notes and avarohana is descending scale of notes. Tala refers to rhythmical cycle, which is like measure in western music. The swaras Sa Re Ga Ma Pa Da Ni correspond to C D E F G A B in western music.

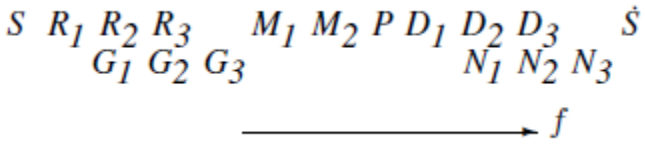

Figure 1. Frequency scale of swaras

Fig 1. shows swaras and in the increasing order of their frequencies of Carnatic music. Some of the swaras has different variations of it. E.g. Ga has three variations Ga1, Ga2 and Ga3. It is also important to note that Ri2 and Ga1 lie in the same frequency scale, so they cannot occur together in a raga. Every song is based out of a base pitch value called Sruti.

Raga identification is essential because raga can be associated with mood and emotion. Some ragas are also found to have therapeutic effects in our body [19][20]. Raga identification requires in-depth knowledge in the raga and precision can be achieved through practice and experience. There are 72 parent ragas which is called Melakarta ragas and thousands of ragas are derived from the Melakarta ragas and they are called Janya ragas. Melakarta ragas are also called Janaka ragas. The melakarta ragas have their arohana strictly in ascending order and their avarohana strictly in descending order. It is also implied that two variations of same swara cannot occur in the same raga.

Gamaka refers to the embellishment of swara (note) where the note is sung with oscillations imposed on it and the transition from one note to another is also sung with oscillations. Raga with same swaras can be distinguished by gamakas. There are a variety of such movements and it varies from one raga to another and the emotion we are trying to express in the song. The way the emotion in the song is expressed is called bhava. These complexities present in Carnatic music makes it difficult to automatically represent Carnatic music.[22][23][27]

Raga identification is performed in Hindustani classical music as well. Hindustani music is another classical music forms which is practiced more on the northern India. Hindustani music also has the similar components of Carnatic music but is different in its own way. The common components between Carnatic music and Hindustani music are raaga, tala, sruti, bhava and the

swaras. Hindustani music also has the concept of arohana and avarohana which refers to increasing and decreasing notes of a raga. Though Hindustani music has the concept of raga, raga identification is different when compared to Carnatic music. The names of ragas are different and classification is done based on 10 thaats and 32 Ragang ragas.The gamakas are more emphasized in Carnatic music than in Hindustani music. The lyrics are very important in Carnatic music but in Hindustani music, it does not matter much. And Carnatic music is more associated to religious aspects whereas Hindustani music does not. The supporting instruments also vary. Vocal Carnatic music concerts are accompanied by violin, mridangam and kanjira whereas Hindustani concerts are accompanied by harmonium and tabla.

Fig1 shows a Carnatic music concert [31]

Figure 2 Carnatic Music Concert

Fig2. Shows a glimpse of Hindustani music concert [30]

Figure 3 Hindustani music concert

This paper organized as follows: the next section presents the most recent related works that are followed by a description of the proposed approach. In section IV, experiments are explained and the results are presented and in section V the results are compared to the baseline. Finally, a conclusion is presented in VI.

II. RELATED WORKS

2.1 Music Information Retrieval

Music Information Retrieval (MIR) is the discipline of retrieving meaningful information from music. The important tasks in the field of MIR are extracting and inferring relevant features from audio signal, symbolically representing music, indexing music using derived features, organizing it in the database, and developing search and retrieval techniques. MIR aims at making music available to listeners based on their interests [3]. Zhouyu et. al emphasize that MIR is comparatively young and started developing in the past decade, given that music is prevailing in our society for a very long period [2]. However, it is seeing a lot of developments in the recent past. The major reasons for rapid increase in technological developments in the field of music are as follows: the development of computing power in personal computers, vast availability of music players including mobile phones and mp3 players and arrival of music streaming platforms like spotify and saavn [5].

In the early stages, MIR research focused on illustrating music in a structured way, which includes representing music in digital format like midi, au, mp3 and wav. Markus et. al reason how the research on MIR started progressing. As per Markus et. al, as digitization became prominent, various signal processing techniques were introduced, which helped in deriving various music qualities like rhythm, timbre, melody and harmony [4]. The features like genre are associated not only to the music content, but also to the cultural aspects that can be generated from a subject matter expert or user interpreted information available in the internet. In mid-2000, music service providers started adding tags for categorization because these tags were required to visualize and organize music collections and generate playlists. As per Lopez et. al, a music piece was treated as a sound wave and system-centric features were derived from it to

categorize it based on a musical or cultural aspect in the initial stages. Recent research in MIR is making a shift by moving away from system-centric features to user-centric features. Casey et al. [5] insist the importance, serendipity and time-awareness of user-centric features which is widely used in recommendation systems these days. For the purpose of evaluation, user-centric features take various music qualities into consideration and human agreement to relationship between the music pieces.

2.2 Music Similarity Estimation

As people are more exposed to digital music, the necessity of recommendation and categorization of music has grown to a great extent as personalization is involved in digitized systems. Music genre classification is one such classification method which has seen a lot of research interest and development in the past decade.

When building a music genre classification method, researchers including, Fu et al [2], use three common descriptors: low-level features, mid-level features and high-level content descriptors. Low-level features are obtained by analyzing the signal obtained by audio waves of a music piece. These features are currently used for various applications but cannot be interpreted by a listener. Examples for low-level features are FFT and MFCC. Mid-level features are again obtained by audio signal but are closer to the attributes understandable to listeners. This include rhythm, contour, interval, harmony and duration. The high-level content descriptors are the sematic labels that provide information about how listeners categorize music. This includes genre, mood, instrument and artist.

Most of the previous research used low-level features as they are not difficult to extract, pre-process, transform and feed it to a classification mechanism. Zheng et. al state that this could run faster but the results produced does not give a logical relationship with the inputs [1]. Fujinaga

and McKay, and Lopez et. al suggest approaches to overcome this limitation and perform genre classification in a musicological way. The first work on analyzing music using mid-level features started in 2004 by Fujinaga and McKay. Fujinaga and McKay worked on analyzing statistical distributions of feature set and applying various machine learning algorithms on it. During the research, they introduced a tool called jMIR, which has in-built functions to perform feature extraction and machine learning from a music file [9].

Zheng et al. [1] proposes a solution which extracts sequential features from symbolic music representations. Melody and bass features of pitch, midi values, duration, contour and interval are chosen. To obtain melody features, the highest music stream is parsed and to obtain bass features, the lowest music stream is parsed. Each feature element is considered as a word and text-based technique called n-gram is performed. Based on the value of n given, word frequency count vector is derived. Multinomial Naïve Bayes classification is performed with word frequency count vector as the input.

Music annotation can be stated as a multi-class classification problem, as we classify it based on one or more attributes. The purpose of annotation of music is to give semantically meaningful tags which could be genre, mood, danceability or style of music. This could be considered as a multi-label learning problem. The classifiers which are used for text-based data (converting music signals to notes) involve pattern recognition. K-Nearest Neighbor (KNN), Gaussian Mixture Model (GMM), Support Vector Machines (SVM), and multinomial naïve Bayes are best suited for finding patterns in text based data [2]. Though Lopes et al. suggested the idea of using text-based features for classification and Zheng et. al implemented it by converting the musicological features to text and used text-based classification for music genre classification

[1][3]. This is a fairly a new idea and could prove to be useful. KNN and SVM are suited for single vector representation and pairwise vector representation.

Apart from the standard machine learning algorithms, neural networks are also often used for MIR problems. Convolutional neural network (CNN) is one such technique which is obtained by taking convolutions over the portions of the input signal of a standard neural network model. CNN can be directly used for classification using feature set. Therefore, CNN can be used for audio classification based on low-level as well as mid-level features. This is exhibited in [7] by applying a convolutional deep belief network (CDBN), which is a type of CNN having multiple layers. Lee et al. state that a specific algorithm cannot be fixed to be used for music classification problems.

In Indian classical music, the initial step involves extracting swaras from the melodic stream where shadja, the base note is identified and the relative notes to it derived based on the shadja. Ranjani et al [22] has proposed a technique to extract shadja from pitch values where shadja is decided based on the pitch values present in various windows in the music file. Based on the shadja, the other notes are calculated and raga prediction is done. Geetha et. al [27] has also worked on extracting swaras using segmentation algorithm and prediction by using string matching algorithm.

Vijay et. al [23] has proposed a technique where pitch values are extracted from the melodic stream, n-gram pitch histograms are obtained and prediction is done using pitch histograms.

On the other hand, using low-level features for raga identification has been researched by Srinath et. al [25]. In their research frequencies are converted to specmurt which is a fast Fourier transformation of the linearly scaled spectrum. Shadja is then found based on the specmurt and based on shadja the other swaras are obtained. The prediction is done using hidden markov

models and the system is proved to perform well for melakarta ragas. The same system does not work well for janya ragas.

Anita et. al [34] have performed classification of ragas using neural networks. Spectral, timbre and tonal features are extracted and classification is performed with back propagation neural network and counter propagation network. The experiments are restricted to melakarta ragas.

III. DATA PREPARATION

3.1.Dataset

In this section, we first describe dataset used for our experiment and the operations performed on the dataset which constituted midi and mp3 files.

3.1.1 Dataset for western music

The dataset used for experiments is a set of 476 files picked from 130000_Pop_Rock_Classical_Videogame_EDM_MIDI_Archive dataset. The files constituted of three genres – Jazz, metal and Ragtime. The dataset contains piano and guitar files and had a mix of various artists. Due the difficulty in using a universal/commonly used MIDI dataset, we have handpicked the files and used it in our experiments. MIDI format files are chosen because it records the musicological aspects of a music which includes notes of each instrument, type of instrument, loudness, finishes, pitch, etc.

Table 1 contains the genres used and the number of files used in each genre.

TABLE 1. DATASET – WESTERN MUSIC

Genre	Number of Records
Jazz	244
Metal	118
Ragtime	114

3.1.2 Dataset for Carnatic music

The dataset used for experiments contains mp3 files from three different ragas taken from dunyacorpmusic[21]. The mp3 files are converted to midi using [24]. The bpm values used for the experiments is 120. The shadja and swaras in Carnatic music is extracted using music 21. Apart from the extracting the swaras, we also extract the duration a swara lasts. The ragas used for experiments are listed in table 2. Out of the three ragas, sankarabharanam is melakarta raga and sindhubhairavi and mohanam are janya ragas.

TABLE 2. DATASET – CARNATIC MUSIC

Raga	Number of Records
Sankarabharanam	19
Mohanam	23
Sindhubhairavi	26

The midi files are read as data streams using music21 python package and labelled with genre name. The highest and lowest score of stream extracted from each music file is used to extract melodic and bass features. The features are then converted to text and n-gram technique is applied on the text features. Classification techniques are applied on n-gram count vector and the genre is predicted.

3.2 Data Pre-processing

In the data pre-processing step, we derive the melodic and bass features from the western music dataset, convert them to text, apply rules to handle special characters and write the data to a CSV file. Fig 4. shows diagrammatic representation of data flow in the pre-processing phase.

In case of Carnatic music, we derive the melodic attributes, convert the pitch, duration, midi, pitch contour, interval and rest duration to text. The rest of the features which are midi counts and note counts remain as numeric in the data frame. We then combine both text and numeric features and write it to a csv file. In the next step where we derive n-grams, we separate text features and apply n-grams only on text features.

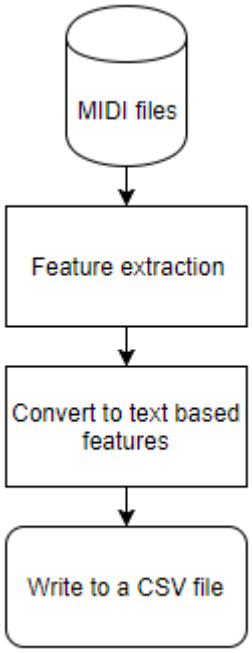

Figure 4 Pre-processing data flow pipeline

Fig 5. Shows the pre-processing data flow pipeline of Carnatic music. The difference in preprocessing is that the count features are separated and only sequential features are converted to text. After the conversion, the two data frames are concatenated and the entire contents is fed to csv file.

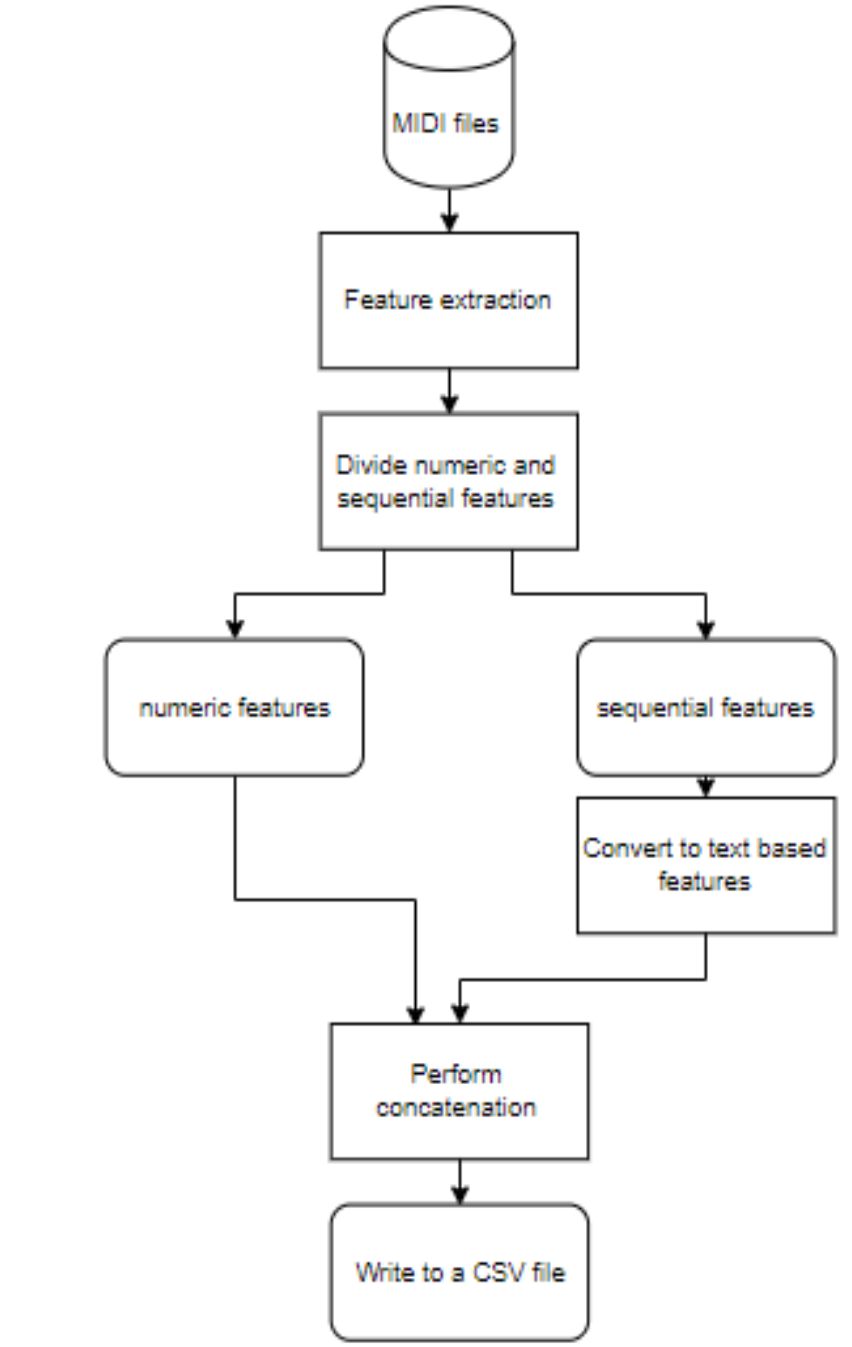

Figure 5. Preprocessing data flow pipeline of Carnatic music

3.2.1 Feature Extraction

3.2.1.1 Western Music Dataset

We are proposing an approach to extract mid-level features like pitch, midi, duration, interval, contour and rest duration of both melody and bass. The features we extract here are short-term features as the live for a very short time. From these short-term features we try to understand the long-term effect of the music piece. Melody features are obtained by parsing highest notes and bass features are obtained by parsing lowest notes.

The features extracted are pitch, midi, duration, contour, interval and rest duration. Pitch is the property which refers to highness or lowness of note. There could be duplicate pitch names like Eb5 and D#5. We are treating them differently as appearing in the midi file. Each pitch has numerical value associated with it and that is the midi value. Midi value ranges from 0 to 127. Higher the value of midi, higher is the pitch. Duration refers to the time which a note last and it is denoted by quarter length. This feature is to understand the tempo of a song. Duration is expressed in quarter length. Contour is a signed integer that represents the difference between two consecutive midi values. It is to denote whether the notes are increasing or decreasing. Contour is obtained by subtracting midi value of previous note from that of the current note. Interval refers to the name of pitch interval between two consecutive notes. Interval is derived by combining previous and current notes and obtaining their pitch interval. Rest refers to the interval of silence in the music stream and is expressed as “rest” to denote the pause. Rest duration is the time taken by a pause and is measured in quarter length. These two attributes are used to understand the tempo of the song.

The above attributes are calculated for both melody and bass streams. All the above attributes are extracted and stored in a panda data frame. The values in each of the fields is converted to text.

The special characters and space in each of the values is replaced with underscore to eliminate discrepancies. The final data frame obtained is stored in a CSV file.

3.2.1.2 Carnatic Music Dataset

Raga classification is done based on the melodic notes of the song. So instead of extracting the features for both bass and melodic stream, the features are extracted only from melodic stream. Apart from extracting the features mentioned in the previous section, count of midi values and note values are also calculated. This is because unlike western music, raga is determined by melodic notes of the song. Emphasis is given to note values because each raga has its own melodic flow of notes. The sequence of notes is spread across the song and it is challenging even for an experienced musician to find which part of the song has the notes identifying the raga. To overcome this difficulty, we compute different values of n-grams of note and pitch sequences and the duration which each of it occurs. We also calculate the count of midi values and note values as songs of same raga are found to revolve around those pitch values and notes. Fig. 6 shows a histogram of pitch values of two songs having sankarabharanam raga.

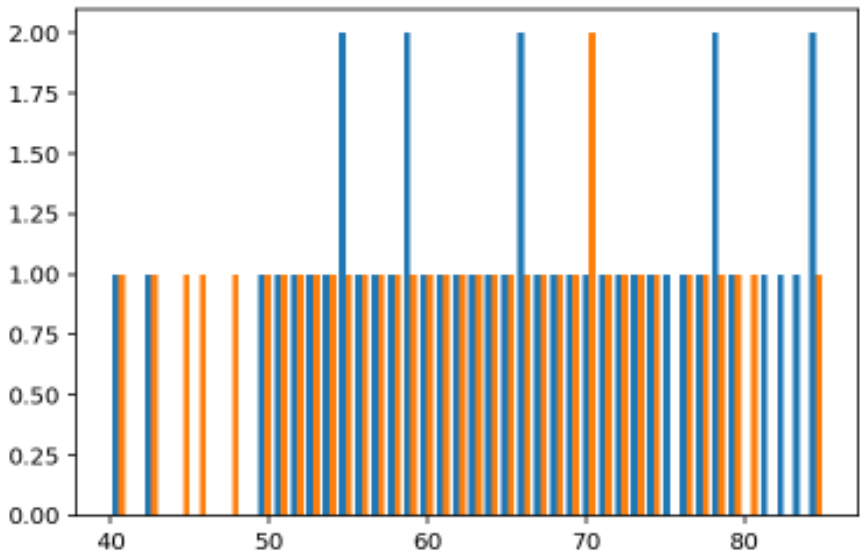

Figure 6. Midi value counts of two songs of the same raga

Fig. 7 shows the note pattern notes from sankarabharanam and mohanam ragas.

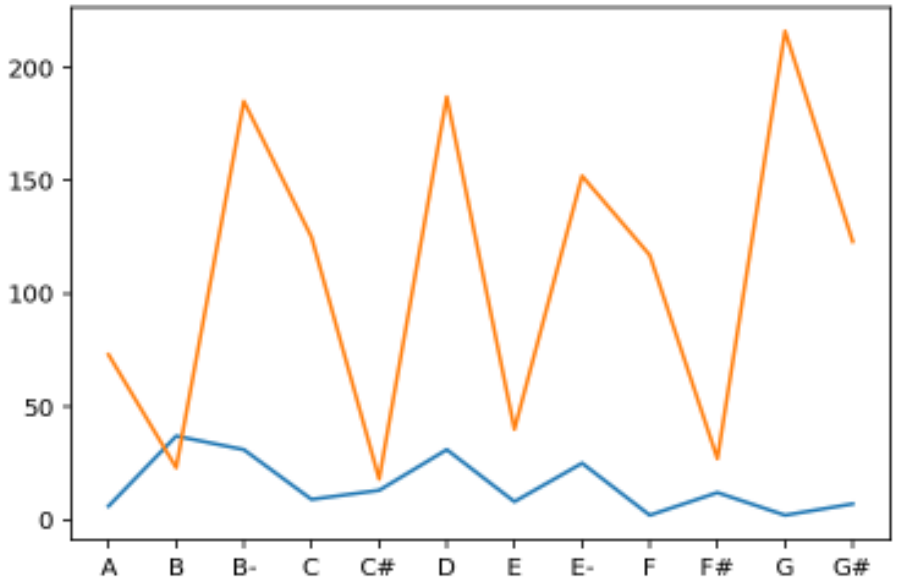

Figure 7. Note pattern counts of two different ragas

3.2.2 N- gram count vector

N gram refers to continuous sequence of n items in given text. We use CountVectorizer function to calculate n-gram matrix. We specify the values as (2,2) and (1,3) for 2-gram and 1,2 and 3-gram vectors respectively. We conducted our experiments for various values of n-gram and maximum accuracy is found to be achieved with 2,3-gram in western music. In case of Carnatic music, the maximum accuracy is achieved using 4,6 grams because a raga's arohana and avarohana has five to eight notes.

3.3 Classification pipeline

After pre-processing, n-gram count vectors are obtained and 5-fold cross-validation is applied. The accuracy is averaged. Data is shuffled while dividing data for training and testing. Train data has 80% of the total set and test data has 20% of the total set. We use Multinomial Naïve Bayes and Random forest classification algorithms for classification. Multinomial Naïve Bayes is used because it works well with text data where classification depends on repetition of multiple words. Random forest is an ensemble classification method which performs classification based on prediction from decision trees and predicts based on maximum voting. Fig 8. Shows the diagrammatic representation of N-gram count vector and classification pipeline.

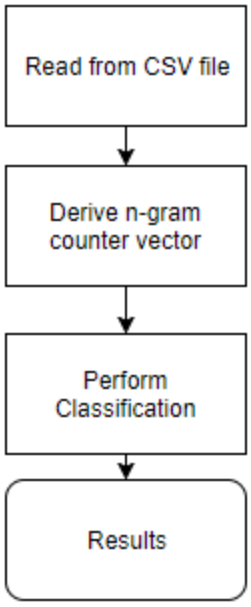

Figure 8. Classification Pipeline data flow

The classification pipeline for Carnatic music slightly varies when compared to western music. The diagrammatic representation is in Fig. 8. The columns in the csv file is divided into text and numeric attributes and n-gram count vectorizer is applied only on text based columns. The other columns are combined to counter vectorized text columns. The data is shuffled and divided into test and train sets in the ration 80:20. Dimensionality reduction is performed using SelectKBest function from scikit learn and the number of columns involved in training and prediction is reduced to 6932 columns. The number 6932 is decided by applying random values of *k*. After dimensionality reduction, the model is trained and prediction is performed. The process is repeated five times for each splitin test sets. The result is the average of five different accuracy values obtained. Multinomial Naïve Bayes and Random forest are the classification algorithms used on preprocessed dataset. Apart from the two algorithms we already used for western music dataset, we also performed classification using neural networks. There are 3 layers in the neural

network. Sigmoid function is applied or the first layer, relu function is applied on the second layer and softmax function is applied on the third layer. The error is measured by ‘categorical_crossentropy. The Sigmoid function is defined as a real-valued, differentiable and a monotonic function to introduce non-linearity. The Relu is defined as maximum of the given value and 0. The Softmax function is usually used in final layer of neural networks for multiclass classification. Dropout is used as the regularization method and is applied in layer 1 and layer 2. The maximum accuracy achieved using neural networks is 80.04% and it is achieved with n-gram (4,4).

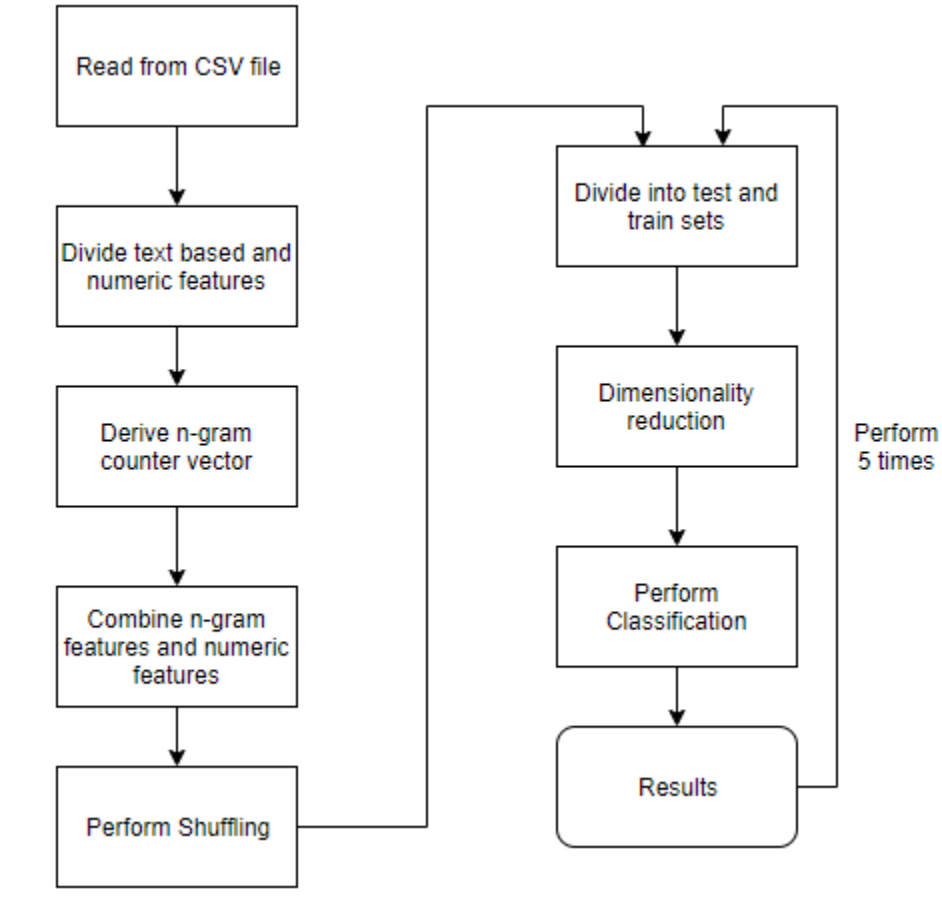

Figure 9. Classification Pipeline data flow for Carnatic music

Python package music21 is used to derive musicological features and scikit-learn, pandas, scipy and numpy packages are used for data pre-processing, prediction and results evaluation. Keras is used for building neural networks. The libraries used are free and open source. We evaluate our results using accuracy. Accuracy is defined as the ratio of number of music pieces identified correctly by the classification algorithm to the total number of music pieces fed to the model for testing.

IV. EXPERIMENTS AND RESULTS

The data from csv file is read and classification is performed for various values of n-gram.

4.1.Western Music

For each value of n-gram the prediction is performed and results are evaluated. The values of n-grams used in the experiments are (1,1), (1,2), (1,3), (2,2) and (3,3). Table 3 provides details about the results obtained by various n-gram values.

TABLE 3. RESULTS EVALUTION – WESTERN MUSIC

N-Gram	Multinomial Naïve Bayes	Random Forest
(1,1)	86	94
(1,2)	84	93
(2,2)	82	94
(1,3)	84	92
(2,3)	88	94
(3,3)	87	95

4.2.Carnatic music

The experiments of Carnatic music are conducted for a wide range of n-grams ranging from (1,1) to (8,8). This is because arohana and avarohana has around 8 notes in it and having a n-gram vector upto length 8 would be useful. Table 4. shows the top 6 results obtained for the exhaustive values of n-grams tested for midi extracted with 120 bpm value.

TABLE 4. RESULTS EVALUTION – CARNATIC MUSIC – 120 BPM

n-gram	Multinomial Naïve Bayes	Random Forest
4,8	84.5	71
4,6	90.375	72.1
5,8	89.85	64.9
5,5	90.05	78.1
2,8	88.025	75.1
4,5	86.25	76.5

Best results are obtained with values of n-grams as 4,6. Table 5. shows the top 6 results obtained for n-grams from (1,1) to (8,8) tested for midi extracted with 180 bpm value.

TABLE 5. RESULTS EVALUTION – CARNATIC MUSIC – 180 BPM

n-gram	Multinomial Naïve Bayes	Random Forest
2,5	86.4	79
4,5	90	77.1
3,6	83.4	81.6
5,7	84.3	79.1
6,7	85.4	71.1
7,7	88.2	86.5

The results of 180 bpm are found to be slightly lesser than results of 120 bpm, highest being 90% using n-gram (4,5).

V. RESULTS

From the results obtained for western music, the best results for genre classification is obtained with (3,3) using random forest algorithm. Other values of n-gram have consistent values with an average of 93%. Fig. 10. gives a comparison of accuracy obtained using approach [1] and proposed approach. When we compare the results, we see that the proposed approach gives consistent results compared to [1].

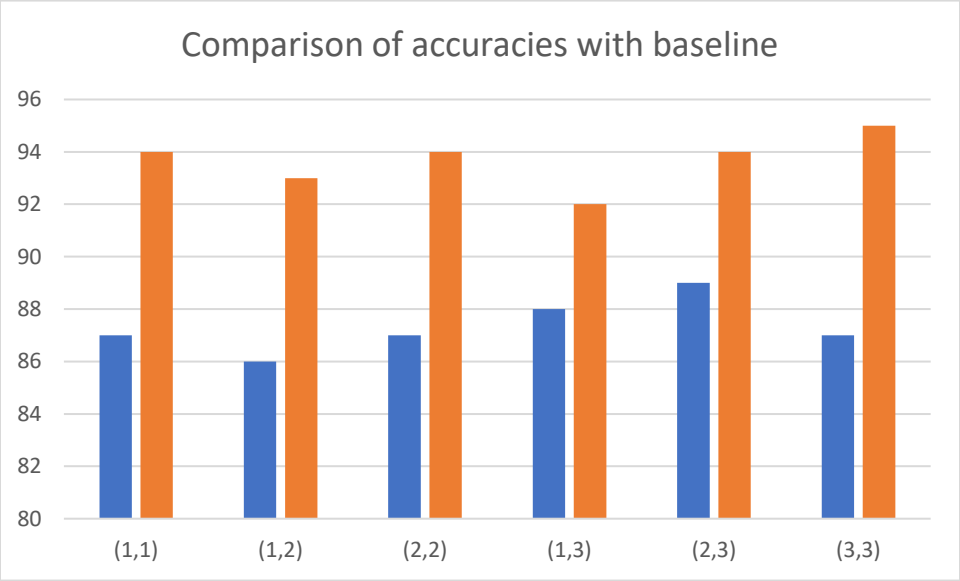

Figure 10. Comparison of Accuracy [1] vs proposed approach

From the comparison we can clearly see that the accuracy has been constantly increased when compared to [1]. Including new features and performing classification using random forest algorithm has improved the prediction of genre when compared to using multinomial naïve Bayes algorithm.

On the contrary, with Indian classical music, the best results are obtained using Multinomial Naïve Bayes algorithm. This is because, Multinomial naïve Bayes works well when data with same class values have repeated patterns of text values. Here in our case, the arohana and the

avarohana of music pieces having same raga files is supposed to repeat. The values of n-gram are also much higher when compared to western music because Carnatic music raga has same pattern of notes in an octave and that is the reason for testing values of n-grams upto 8. The length of arohana and avarohana is usually of length 8 but due to the improvisations done by musicians, the exact arohana and avarohana does not appear in the extracted pitches. The best results with the value of 90.3 is obtained for n-gram (4,6) for bpm value of 120. This implies that notes of 4 to 6 length appear in the music file following the pattern of the raga. The baseline results for Carnatic music dataset is obtained from [22] and the best achieved by baseline is 84.6% using 4-gram. The results obtained by the proposed approach shows an improvement of 6%. The average accuracy values are also found to be greater than the baseline.

Fig 11 shows the comparison of accuracies of midi values extracted with bpm value of 120. Multinomial Naïve Bayes is found to perform consistently better when compared to Random forest algorithm which proves our theory.

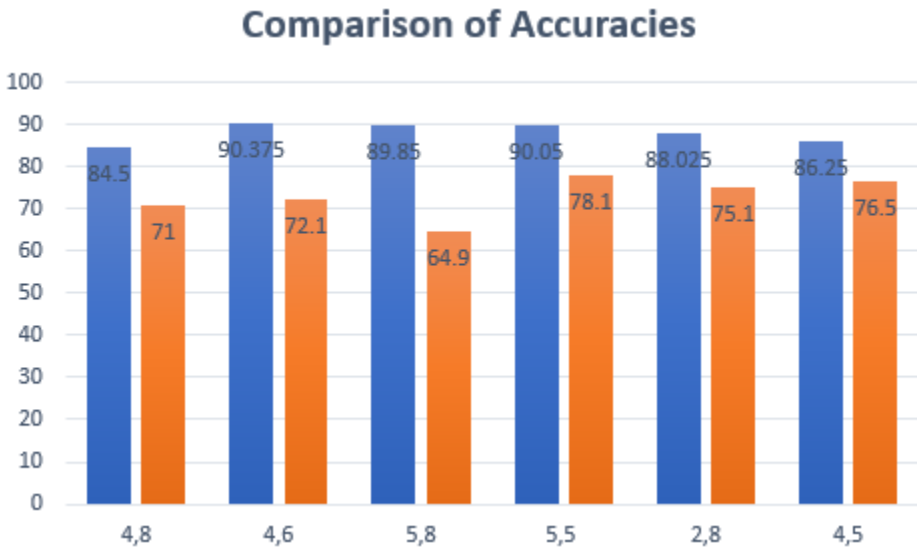

Figure 11. Comparison of accuracies of Multinomial Naïve Bayes and Random Forest algorithm of files with 120 bpm

Fig 11 shows the comparison of accuracies of midi values extracted with bpm value of 180. Multinomial Naïve Bayes is found to perform consistently better when compared to Random forest for bpm values of 180 as well.

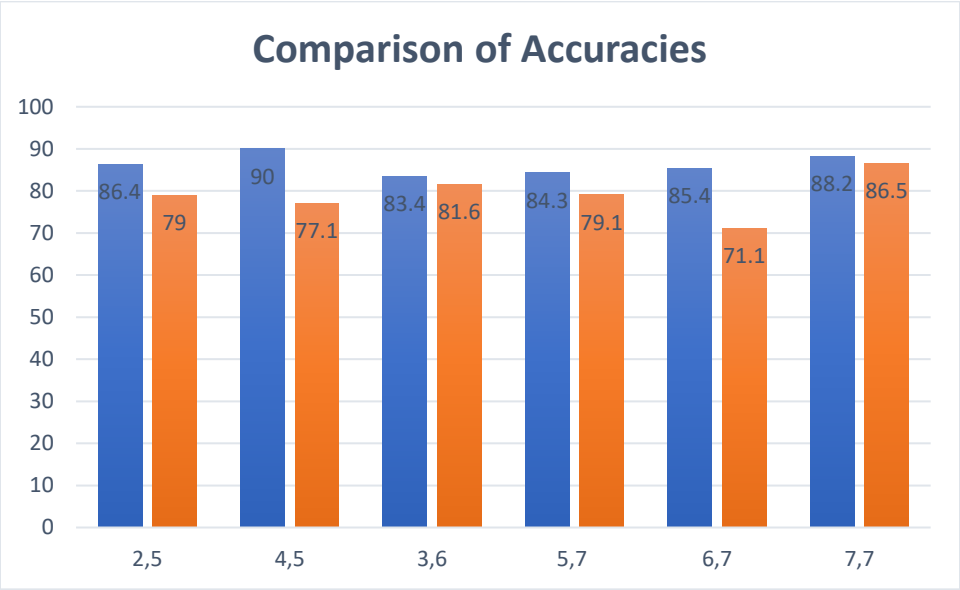

Figure 12. Comparison of accuracies of Multinomial Naïve Bayes and Random Forest algorithm of files with 120 bpm

VI. CONCLUSION AND FUTURE WORK

With the advancement in the field of MIR, music similarity estimation in a musicological way has proved to provide better solutions for many MIR problems such genre classification and emotion classification [1]. The research of deriving high-level features from musicological features can be compared with the results of using low-level features and the gaps can be filled [3]. The research validates that the approach of using musicological features can overcome limitation of using low-level features and build a solution which could be logically related to features understood by listeners.

The research has demonstrated that the approach of using mid-level features has proven its concept, and has a promising future for further developments. This includes extracting, computing and combining more features, extracting and preparing more musical elements, using larger, comprehensive, and various instrumental music datasets and incorporating other advanced text classification techniques. The techniques discussed in the literature survey is not only useful for genre classification, but can also be extended to other classification systems like emotion classification and query by humming problems.

On the other hand, for Indian classical Carnatic music previous research has focused mainly on mid-level features as the raga can only be understood by its swara placements. But, in the proposed approach we have explored the idea of using midi counts and note counts. This could be further improvised by using the duration a swara. For every n-grams, instead of incrementing the count the duration for which the swara pattern lives can be incremented. This would help us to identify the movement of swaras which varies from one raga to another raga. We have seen that using mid-level features in Carnatic music has proved its concept. We can also experiment a combination of low-level and mid-level features and use it for prediction. The current research is

limited to the compmusic dataset and training has been done only with less than hundred files. The train data could be expanded or the same experiments can be conducted with a different dataset.

VII. REFERENCES

[1] E. Zheng, M. Moh, and T. Moh, "Music genre classification: A n-gram based musicological approach," in *the Proc. of IEEE 7th Int. Advance Computing Conf.,* 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7976875/

[2] Z. Fu, G. Lu, K. Ting and D. Zhang, "A Survey of Audio-Based Music Classification and Annotation," in *the Proc. of IEEE Transactions, 2011.*

[3] T. Huang, and P. Chang, "Large-scale cover song retrieval system developed using machine learning approaches," in the *Proc. of IEEE Int. Symp. on Multimedia,* 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7823695/

[4] A. T. Lopez, E. Aguiar, and T.O. Santos, "A comparative study of classifiers for music genre classification based on feature extractors," in *the Proc. Of Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER),* 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7806258/

[5] M. Schedl, E. Gomez, J. Urbano, "Music information retrieval: recent developments and applications," in Music Information Retrieval vol.8, no 2-3, pp 127-261, Jan 2014. [Online]. Available: http://www.nowpublishers.com/article/Details/INR-042

[6] A. R. Rajanna et. al, "Deep neural networks: a case study for music genre classification," in *Proc. Machine Learning and Applications (ICMLA), IEEE 14th Int. Conf,* 2016 [Online]. Available: http://ieeexplore.ieee.org/document/7424393/

[7] http://cmc.music.columbia.edu/MusicAndComputers/chapter2/02_01.php

[8] H. Lee, Y. Largman, P. Pham, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Proc. Advances in Neural Information Processing Systems,* 2009.

[9] M. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval: Current directions and future challenges," in *Proc. IEEE, vol. 96, no. 4, pp. 668–696,* 2008.

[10] C. McKay and I. Fujinaga, "Automatic genre classification using large high-level musical feature sets." in *ISMIR,* vol. 2004, 2004, pp. 525– 530.

[11] https://www.nap.edu/read/6323/chapter/11

[12] E. Pampalk, T. Pohle, and G. Widmer, "Dynamic playlist generation based on skipping behaviour," In the Proc. of *ISMIR 2005 Sixth Int. Conf. on Music Information Retrieval*, September 2005.

[13] K. West and S. Cox, "Finding an optimal segmentation for audio genre classification," *In the Proc. ofISMIR 2005 6th Int. Conf. on Music Information Retrieval,* September 2005.

[14] scikit-learn: machine learning in python. [Online]. Available: http://scikit-learn.org/

[15] music21: a toolkit for computer-aided musicology. [Online]. Available: http://web.mit.edu/music21/

[16] C. D. Manning, P. Raghavan, and H. Schutze, Introduction to information retrieval. Cambridge university press Cambridge, 2008, vol. 1.

[17] R. Hillewaere, B. Manderick, and D. Conklin, "String methods for folk tune genre classification." *in the Proc.* ISMIR, vol. 2012, 2012, p. 13th.

[18] J. Jakubik H. Kwaśnicka, "Music emotion analysis using semantic embedding recurrent neural networks," in *Proc. Innovations in Intelligent Systems and Applications (INISTA), 2017 IEEE International Conf*[Online]. Available: http://ieeexplore.ieee.org/document/8001169/

[19] http://pilu.in/raga-therapy.html

[20] https://ayurveda-foryou.com/music/raga_chikitsa.html

[21] A. Srinivasamurthy, & X. Serra. “A Supervised Approach to Hierarchical Metrical Cycle Tracking from Audio Music Recordings”. *In Proceedings of the 39th IEEE International Conference on Acoustics,* Speech and Signal Processing (ICASSP 2014) (pp. 5237–5241). Florence, Italy.

[22] H. G. Ranjani, G. Arthi, T.V. Sreenivas,” Carnatic music analysis: shadja, swara identification and raga verification in alapana using stochastic models,” *In Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011*.

[23] V. Kumar, H. Pandya and C. V. Jawahar, “Identifying Ragas in Indian music,” *In Proc. of International Conference on Pattern Recognition*), 2014

[24] https://github.com/justinsalamon/audio_to_midi_melodia

[25] http://essentia.upf.edu/

[26] https://librosa.github.io/librosa/

[27] N. A. Jairazbhoy, “C. SubrahmanyaAyyar: Acoustics for music students”,[viii], 71 pp., 3 plates., *Bulletin of the School of Oriental and African Studies*, vol. 24, no. 01, p. 71, Dec. 2009.R. Sridhar and T. V. Geetha, Swara identification of Carnatic music, *in Proc. IEEE Computer Society Press, Proceeding of ICIT 2006*

[28] K. R. Scherer M. Zentner "Emotion effects of music: Production rules" in Music and emotion: Theory and research Oxford University Press pp. 361-392 2001.

[29] S. Sigtia S. Dixon "Improved music feature learning with deep neural networks" *in Proc. of the 38th International Conference on Acoustics Speech and Signal Processing* (ICASSP) pp. 6959-6963 2014.

[30] http://harmonyom.blogspot.com/2012/05/

[31] https://www.youtube.com/

[32] M. Henaff K. Jarrett K. Kavukcuoglu Y. LeCun "Unsupervised learning of sparse features for scalable audio classification" ISMIR vol. 11 no. 445 2011

[33] http://harmonyom.blogspot.com/2012/05/

[34] N. Glazyrin "Mid-level features for audio chord recognition using a deep neural network" Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki vol. 155 no. 4 pp. 109-117 2013.

[35] R. Anita, K. Gunavathi, A. Asokan, “Classification of Melakartha Ragas Using Neural Networks, *“in the proc. 2017 International Conference on Innovations in Information, Embedded and Communication Systems*