JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

The Effects of a Spiking Neural Network on Indian Classical Music

Dr.D.Shanthi
Associate Professor
Maturi Venkata Subba Rao Engineering College

Abstract: Indian Classical Music is one of the world's oldest music which made wonders in both ancient and modern times. This research shows how ragas and rasas affect human emotions. To demonstrate this, we compared the results of CPNN, BPNN, and PNN with those of Spiking Neural Networks, a newer machine learning technology.

Keywords: Raga, Indian Classical Music, Lyrics, Spiking Neural Network, emotion.

1. Introduction:

The need to analyze massive sets of music recordings automatically has made music information retrieval an important study subject in recent years.

Streaming firms are one example that requires MIR. Detecting mood from music has long been an important field in MIR research. It can automatically discern emotions in lyrics and music.

It focuses on multimodal mood detection using song lyrics and audio signals. The mood of ICM (Indian Classical Music) is difficult to discern. It contains two parts: raga and rasa.

The combination of given basic notes are referred as ragas and they are Sa (Shadja), Ri(Rishabha), Ga (Gandhara),Ma (Madhyama), Pa (Panchama), Dha(Dhaivata), Ni(Nishada)]. There are nine rasa's in ICM and they are 1) Shringar (Love), 2)Hasya(Humor), 3) Karuna(Pathos), 4) Rudra (Anger), 5) Veer (Heroism), 6) Bhayanaka(Terror), 7) Vibhatsa (Disgust), 8) Adbhuta (Wonder), 9)Shanta (Calm).

1.1 Spiking Neural Network

A Spiking Neural Network is a third-generation neural network model, while CPNN, BPNN, and PNN are second-generation neural networks that increase the level of realism in a neural simulation. Spiking neural networks have proven to be powerful functional computational tools. more physiologically plausible and faster than second-generation networks. Applications include function approximation, categorization, and pattern recognition. Signal processing, speech recognition, picture processing, event detection, spatial navigation, or motor control are all possible applications of spiking networks in applied engineering. Spiking neural networks can solve any problems that non-spiking neural networks cannot solve, and are more powerful computationally than perceptrons and sigmoid gestures. For this reason, scientists are extremely involved in spiking neural networks. more adaptable for computer vision applications and with strong biological relevance.

1.2 Back Propagation Neural Network:

In a multilayered feedforward network with a continuous differential function, this learning algorithm also uses a differentiable gradient descent. Gradient descent reduces errors to 0 (or less).

Propagation of an error to a hidden unit, it is not only back propagated to the output layer, but also to the hidden layer. BPNN training is a three-stage process. Feed forwarding, error propagation, and weight update

1.3 Counter Propagation Neural Network:

A hybrid network combines the best of two or more network designs suggested by Hecht-Nielsen 1986.

The hidden layer is an unsupervised Kohonen network, and the output layer is a Grossberg (outstar) network. The Widrow-Hoff rule trains the output layer. It allows pattern output rather than just a category number can be seen as a two-way associative memory.

1.4 Probabilistic Neural Network (PNN):

A probabilistic neural network model has three layers: input, hidden, and output. The network is trained using a gradient-descent technique. For each instance t, the desired output is a 1 for the left child and a 0 for the right child. Formula y_t yields the true output:

2. Literature Survey

Music affects our emotions, and music has varying effects on our brains. Music is defined as a structured sound that resonates with nerve tissue [1]. Babies who don't understand music yet respond to what they hear. It is a biological sense where rhythm and brain interact [2,3]. Many scholars have studied the relationship between emotion and music [4].

Barthet et al.[5] defined MER as music emotion recognition. According to Wieczorkouska et al.[6], Many researchers later agreed with him. In this system, each musical section is allocated various labels based on its characteristics. It represents an emotion. Extracting and selecting features for music classification is a difficult operation [7].

Acoustic characteristics are the most commonly used [8]. Timbre, spectral and rhythmic properties. Rhythmic characteristics are extracted from a histogramic beat [9]. Timbre features include MFCC and chroma [11]. Spectral properties include centroid, flatness, flux, rolloff, and crest factor.

Despite several improvements, it is still not sufficient. Getting the exact key acoustic elements that genuinely effect our emotions is tough due to the numerous inconsistencies of feelings produced by similar music frames or portions for us.

Emotion classification in Indian music is a new task. Indian classical music creation has a profound impact on the human mind. ICM has two parts: raga and rasa. These notes are called raga swaras and are spelled as [Sa, Ri, Ga, Ma, Madhyama, Pa, Dha, Ni].

Every raga has notes and a mood. There are nine rasas: Shringar (Love), Hasya (Humor), Rudra (Anger), Veer (Heroism), Bhayanaka (Terror), Adbhuta (Wonder), and Shanta (Wonder) (Calm).

To assess a piece of music, one must first identify the raga's characteristics that contribute to the mood. Spiking neural network is an advanced machine learning method used for emotion recognition. It can discern emotion in text and voice as well as music. SNN is a synthetic neural network that closely resembles natural neural networks. SNN has operational model, neural, and synaptic state notions. Unlike multilayer perception networks, SNNs fire only when the membrane potential, or internal quality of the neuron, reaches a predefined value.

When a neuron fires, it sends out a signal to its neighboring neurons, though it affects its potential. In SNN, the current activation is treated as a differential equation, where the entering spikes push with a higher value or decaying.

The outgoing spike train can be interpreted as a real number, depending on spike frequencies or the period between spikes.

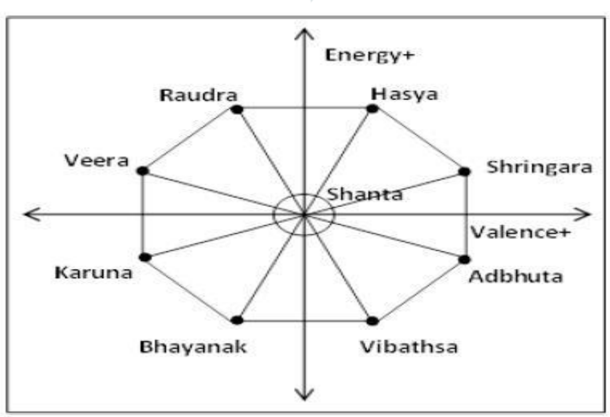
Methodology:

First, we do audio segmentation, also known as framing. The bigger raga or audio tape is broken into multiple fragments, each displaying different rasas. So, pre processing includes converting the song to an audio format. We create wave files from various audio clips. The waves become spike trains. For spectral analysis of frames, we employ STFT. Then spectral, temporal, and energy features are extracted. Spectral properties include spectral roll-off, flux, centroid, and Mel Frequency Cepstral Coeffients (MFCC).

Features of zero crossing rate and compactness The energy-related property is Root Mean Square (RMS). Training the feature vectors with the requisite feature set and activation weights produces random stochastic values.

During classification, the classifier model is trained and tested against an existing dataset of spikes based on rasas for various ragas. Because ICM ragas might have several emotions (rasa), the classification must be multi-label. Our system uses SNN for categorization. Uses a four-layer network with Softmax as the output layer The softmax layer has nine labels to map every conceivable rasa produced from the raga section.

The nine rasas are mapped with Thayers Valence arousal model to obtain emotions. The nine labels map output layers. The first quadrant has Shringara rasa (love) and Hasya rasa (laughing). The fourth quadrant has Adbhuta (surprise) and Vibhasta (disgust). The second quadrant has the words veera and raudra rasa. Karuna rasa denotes sadness while bhayanaka rasa indicates dread, both in the third quadrant. The Shanta Rasa is the state of transcendence, tranquilly, and serenity.



Most audio files in the.wav format are uncompressed LPCM (Linear Pulse Code Modulation), however others are compressed. However, compression might cause data loss. Thus, lossless song format delivers more accurate emotional consequences. Since ICM has no standard database, the database is built from internet raga singers. To expand the dataset volume, each audio clip is split into 10ms chunks, termed data clips.

2. Algorithm:

1) Start
2) Consider a data set d1 consisting of 'n' ragas.
3) Train d1
4) Pre-process the data
5) Acoustic Feature extraction from 'd1' data.
6) The wave format of dataset 'd1' is converted to spikes set 's1'.
7) SNN, a neural network (unsupervised machine learning) used to determine internal data representation of dataset 'd1'.
8) In this case, the rasa must be found for another dataset 'd2.' #d2: test the dataset
9) Process data from 'd2'.
10) Acoustic feature extraction from dataset 'd2'.
11) The wave format of dataset 'd2' is translated into spikes set 's2'.
12) (unsupervised machine learning) SNN used to discover internal data representation of dataset 'd1'.
13) Comparison of spike set 's1' with spike set 's2' and data prediction.
14) Emotion or rasa recognition is evaluated by comparing it to specified values.
15) End

Let us consider a data set d1 which is pre-determined of 'n' number of ragas.

3. Mathematical Expressions:

1) Segment the songs 10ms each
2) Apply STFT(Short Time Fourier Transform), and find the frequency.
3) Extract features like melodious frequency, pitch and zero crossing rate using the given formulae

$$M(f)=1125 \ln(1+f/700)$$

$$\text{Pitch} = \{[(2)*1/1200]^{n}\} *220 \text{ Hz}$$

$$Z=\Sigma|\text{sign}(x(n) - \text{sign}((n-1))|$$

4. Results and Comparisions:

Sl.no	Technique	Results obtained
1	Counter Propagation Neural Network	54%
2	Back Propagation Neural Network	64%
3	Probabilistic Neural Network	68%
4	Spiking Neural Network	72%

According to the results of a literature review, the notion of neural networks has not been employed in this emotional extraction from music. For this reason, we used all second generation and third generation technologies available in Indian classical music and were successful in extracting some of the desired results to a certain extent.

5. Conclusion:

Using all second and third generation neural networks, we found that CPNN provided 54 %, BPNN provided 64 %, and PNN provided 68 %, all of which are better results than using analytical models. However, when we used SNN, we found that it provided outstanding results, i.e. % accuracy. As a result of these findings, we can confidently state that SNN is the most recent concept to provide positive outcomes, and that it can be used to a variety of other domains such as big data and so on.

References:

[1] Kate Hevner. Experimental studies of the el ements of expression in music. The American Journal of Psychology, 48(2):246–268, 1936. 374 Proceedings of the 19th ISMIR Conference, Paris, France, September 23-27, 2018.

[2] Mireille Besson, Frederique Faita, Isabelle Peretz, AM Bonnel, and Jean Requin. Singing in the brain: Independence of lyrics and tunes. Psychological Science, 9(6):494–498, 1998.

[3] Youngmoo E Kim, Erik M Schmidt, Raymond Migneco, Brandon G Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A Speck, and Douglas Turnbull. Music emotion recognition: A state of the art review. In ISMIR, pages 255–266, 2010.

[4] Tao Li and Mitsunori Ogihara. Detecting emotion in music. In ISMIR, pages 239–240. Johns Hopkins University, 2003. [19] Thomas Lidy and Alexander Schindler. Parallel convolutional neural networks for music genre and mood classification. MIREX, 2016.

[5] George Tzanetakis. Marsyas submissions to MIREX 2007. In ISMIR, 2007.

[6] Geoffroy Peeters. A Generic Training and Classification System for MIREX08 Classification Tasks: Audio Music Mood, Audio Genre, Audio Artist and Audio Tag. In MIREX, Philadelphia, United States, September 2008.

[7] Dan Yang and Won-Sook Lee. Disambiguating music emotion using software agents. In ISMIR, volume 4, pages 218–223, 2004.

[8] Shlomo Argamon, Marin Sari ˇ c, and Sterling S Stein. ´ Style mining of electronic messages for multiple authorship discrimination: first results. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 475–480. ACM, 2003.

[9] Cyril Laurier, Jens Grivolla, and Perfecto Herrera. Multimodal music mood classification using audio and lyrics. In Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on, pages 688–693. IEEE, 2008.

[10] Feng Su and Hao Xue. Graph-based multimodal music mood classification in discriminative latent space. In International Conference on Multimedia Modeling, pages 152–163. Springer, 2017.

[11] Xiao Hu and J Stephen Downie. Improving mood classification in music digital libraries by combining lyrics and audio. In Proceedings of the 10th annual joint conference on Digital libraries, pages 159–168. ACM, 2010.

[12] J Stephen Downie. The music information retrieval evaluation exchange (mirex). D-Lib Magazine, 12(12):795–825, 2006.

[13] Thomas Lidy and Alexander Schindler. Parallel convolutional neural networks for music genre and mood classification. MIREX, 2016.

[14] Byungsoo Jeon, Chanju Kim, Adrian Kim, Dongwon Kim, Jangyeon Park, and Jung-Woo Ha. Music emotion recognition via end-to-end multimodal neural networks.

[15] Moyuan Huang, Wenge Rong, Tom Arjannikov, Nan Jiang, and Zhang Xiong. Bi-modal deep boltzmann machine based musical emotion classification. In International Conference on Artificial Neural Networks, pages 199–207. Springer, 2016.