K-Nearest Neighbour and Earth Mover Distance for Raaga Recognition

Prof. Prasad Reddy P.V.G.D
Department of CS&SE, Andhra University
Visakhapatnam, India

D B. Tarakeswara Rao
School of Computing University
Vignan University
Guntur, India

Dr. K.R Sudha
Department of EEE,
Andhra University,
Visakhapatnam, India

Hari CH.V.M.K
Department of IT,
GITAM University,
Visakhapatnam, India

ABSTACT

As far as the Raaga Recognition process, most probably the significant and straightforward classifier in the arsenal or machine learning techniques is the Nearest Neighbour Classifier. The classification is achieved by identifying the nearest neighbours to a query example and using those neighbours to determine the class of the query. This approach to classification is of particular importance today because issues of poor run-time performance are not such a problem these days with the computational power that is available.

This paper presents an overview of techniques for Nearest Neighbour classification focusing on; mechanisms for finding distance between neighbours using Cosine Distance (CD), Earth Movers Distance (EMD) and formulas are used to identify nearest neighbours, algorithm for classification in training and testing for identifying raagas. From the results it is concluded that Earth Movers Distance (EMD) is producing better results than Cosine Distance measure.

Keywords--- Raaga, Cosine Distance(CD), Earth Movers Distance (EMD), K-NN

1. INTRODUCTION AND PROBLEM DEFINITION

The most outstanding two systems of classical music in the Indian subcontinent is a melodic system known as raaga. Performance in Indian classical music is always within a raaga, except for solo percussion. Raaga is a system within which performers improvise and compose. Raagas are often summarized by the notes they use, though many raagas in fact share the same notes. Raaga recognition is a difficult task even for humans. A raaga is popularly defined as a specified combination, decorated with embellishments and graceful consonances of notes within a mode which has the power of evoking a unique feeling distinct from all other joys and sorrows. It possesses something of a transcendental element.

A raaga is characterized by several attributes, like its Vaadi-Samvaadi, Aarohana-Avrohana and Pakad [19], besides the sequence of notes. It is of utmost importance to note here that no two performances of the same raaga, even two performances by the same artist, will be identical. A certain music piece is considered a certain raaga, as long as the attributes associated with it are satisfied. This concept of Indian classical music, in that way, no doubt, is very open.

Given an audio sample, find the underlying raaga for the input, complexity of a raaga is highly variable in performance. Though tried to be very general in the approach, some constraints have to be placed on the input.

Through this work the following major contributions to the study of musical raagas and KNN with CD and EMD are made. In first place, our solutions based primarily on techniques from speech processing and pattern matching, which shows that techniques from other domains can be purposefully extended to solve problems in computational musical raagas, Secondly, the two note transcription methods presented are novel ways to extract notes from sample raagas of Indian classical music. This approach has given very encouraging results.

The rest of the paper is organized as follows. Section 2 highlights some of the useful and related previous research work in the area. The solution strategy is discussed in detail in Section 3. The test procedures and experimental results are presented in Section 4, Finally, Section 5 lists out the conclusions.

2. LITERATURE REVIEW

The work of literature in Carnatic music retrieval is on a slow pace compared to western music. Some work is being done in Swara identification [1] and Singer identification [2] of Carnatic music. In Hindustani music work has been done in identifying the Raaga of Hindustani music [3]. In [3] the authors have created a HMM based on which they have identified two raagas of Hindustani music. The fundamental difference between Hindustani Raaga pattern and Carnatic Raaga pattern is that in Hindustani R1, R2 are present as against R1, R2, R3 in Carnatic. Similarly G, D, N all has three distinct frequencies in Carnatic music as compared to two frequencies in Hindustani [8]. This reduces the confusion in identifying the distinct frequencies in Hindustani music as compared to Carnatic music. The authors have not used polyphonic music signal and have assumed that the input music signal is a voice only signal. The fundamental frequency of the signal was also assumed and based on these features the raaga identification process was done for two Hindustani raagas.

On the western music aspect, melody retrieval is being performed by researchers. The one proposed by [9] is based on identifying the change in frequency in the given query. The query is received in the form a humming tune and based on the rise and fall in the pitch of the received query, the melody pattern that matches with the query's rise and fall of pitch is retrieved. The melody retrieval based on features like distance measures and gestalt principles. The approach is based on low level signal features and the raaga is identified by considering different instrument signal as input to our system. In the present work raaga identification is done using KNN with two different distance metrics one CD and the other EMD.

3. PROPOSED SOLUTION

K Nearest Neighbour has rapidly become one of the booming technologies in today's world for developing convoluted control systems. Melakarta Raaga Recognition is one the fascinating applications of KNN – which is basically used in raaga identification for many cases, raaga detection is considered as a rudimentary nearest neighbour problem. The problem becomes more fascinating because the content is an audio – given an audio find the audio closest to the query from the trained database.

The intuition underlying Nearest Neighbour Classification is quite straight forward, classified based on the class of their nearest neighbours. It is often useful to take more than one neighbour into account so the technique is more commonly referred to as k-Nearest Neighbour (k-NN) Classification where k nearest neighbours are used in determining the class. Since the training examples are needed at run-time, i.e. they need to be in memory at run-time, it is sometimes also called Memory-Based Classification. Because induction is delayed to run time, it is considered a Lazy Learning technique. Because classification is based directly on the training examples it is also called Example-Based Classification or Case-Based Classification.

The basic idea is as shown in Figure 1 which depicts a 3-Nearest Neighbour Classifier on a two-class problem in a two-dimensional feature space. In this example the decision for q1 is straightforward – all three of its nearest neighbours are of class O so it is classified as an O. The situation for q2 is a bit more complicated at it has two neighbours of class X and one of class O. This can be resolved by simple majority voting or by distance weighted voting (see below). So k-NN classification has two stages; the first is the determination of the nearest neighbours and the second is the determination of the class using those neighbours. The following section describes the techniques CD and EMD which is used to raaga classification.

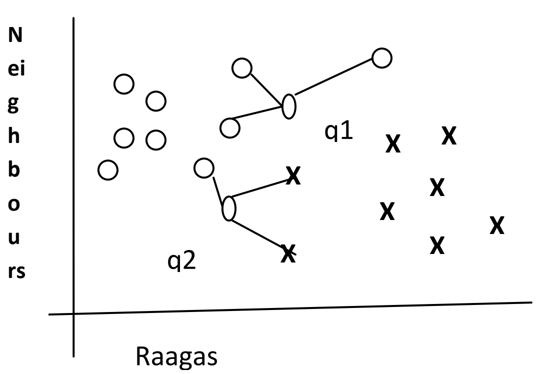

Figure 1: A simple example of 3-Nearest Neighbour Classification

3.1 Cosine Distance and Earth Mover Distance

3.1.1 Cosine Distance

Cosine similarity (CD) between two vectors x and y is defined as:

$CD(x; y) = (x^T * y)/(||x||*||y||)$ ------- (1)

Cosine similarity has a special property that makes it suitable for metric learning: the resulting similarity measure is always within the range of -1 and +1. this property allows the objective function to be simple and effective.

3.1.2 Earth Mover Distance (EMD)

The solution to a discrete optimal mass transportation problem is the EMD. It represents the minimum cost of moving earth from some source locations to fill up holes at some sink locations. In other words, given any two mass (or probability) distributions, one of them can be viewed as a distribution of earth and the other a distribution of holes, then the EMD between the two distributions is the minimum cost of rearranging the mass in one distribution to obtain the other. In the continuous setting, this problem is known as the Monge-Kantorovich optimal mass transfer problem and has been well studied over the past 100 years the importance here is that EMD can be used to measure the discrepancy between two multidimensional distributions.

3.2. Proposed Methodology/Algorithm for Raaga Recognition System

The following is the methodology used for the Melakarta Raaga Recognition for training and testing. Initially first k-Nearest Neighbour Classifier is determined on a two-class problem in a two-dimensional feature space which is shown in the following diagram raagas in horizontal axis and neighbours of raaga on the vertical axis. In this proposed approach the decision for raaga is straightforward – one of its nearest neighbours is of class O and one of class X.

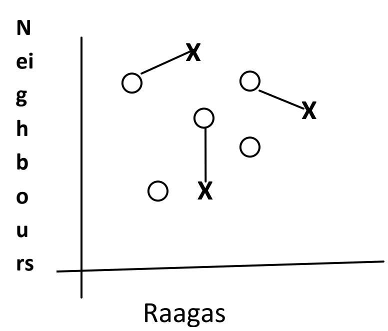

Figure 2: 1-Nearest Neighbour classification of Raagas

A training dataset D is made up of (xi), I Є[1,|D|] training samples where xi is the raaga. The raaga is divided in to 15 samples by eliminating unwanted frequencies (disturbances, accompanied instruments) by using low level filter-Fourier Transform of a Signal (Spft). The same process is repeated for each raaga in database D. Then these samples are trained by using Self- Organizing and Learning Vector Quantization Nets. The grouping process is carried by us. Each training example is labeled with a class label yj Є Y. Our objective is to classify an unknown example raaga q. Now training process is completed. Next the testing phase is performed by using KNN classification.

The KNN approach carried in two phases

1 Determination of Nearest Neighbours
2 Determination of the class using those neighbours

Determination of Nearest Neighbours :

For each xi Є D the distance between q and xi is calculated as follows:

$$d(q,x_i) = \sum_{f \in F} \varpi_f \delta(q_f, x_{if}) \quad \text{------(2)}$$

Where xi = trained raaga ,

q = testing raaga,

f = feature(flow pattern)

wf = weighted feature of raaga

There are huge ranges of possibilities for this distance metric; a basic version for continuous and discrete attributes would be:

$$\delta(q_f, x_{if}) = \begin{cases} 0 & f \textit{ discrete and } q_f = x_{if} \\ 1 & f \textit{ discrete and } q_f \neq x_{if} \\ |q_f - x_{if}| & f \textit{ continuous} \end{cases}$$

----(3)

The k nearest neighbours is selected based on this distance metric. In order to determine the class of q the majority class among the nearest neighbours is assigned to the query. It will often make sense to assign more weight to the nearer neighbours in deciding the class of the query.

Determination of the class using those neighbours:

If more than one of the neighbours is identified then it can be resolved by simple majority voting or by distance weighted voting. A fairly general technique to achieve this is distance weighted voting where the neighbours get to vote on the class of the query case with votes weighted by the inverse of their distance to the query.

$$Vote(y_i) = \sum_{c=1}^{k} \frac{1}{d(q,x_c)^n} 1(y_j, y_c) \quad \text{------ (4)}$$

Thus the vote assigned to class yj by neighbour xc is 1 divided by the distance to that neighbour, i.e. 1(yj , yc) returns 1 if the class labels match and 0 otherwise. From the above equation would normally be 1 but values greater than 1 can be used to further reduce the influence of more distant neighbours. Now the distance measures Cosine and EMD measures applied to our KNN process is discussed.

3.2.1 Cosine Distance Measure

The cosine similarity measure is the cosine of the angle between these two vectors, suppose di and dj are the paths between ai and aj in instance xi and instance xj, respectively. di and dj are represented as vectors of term frequencies in the vector-space model. The cosine is calculated by using the following formula

$$\cos(d_i, d_j) = \frac{\sum_k a_{i,k} \cdot a_{j,k}}{\sqrt{\sum_k a_{i,k}^2}\sqrt{\sum_k a_{j,k}^2}} \quad \text{----- (5)}$$

3.2.2 Earth Mover Distance

The Earth Mover Distance (EMD) is a distance measure that overcomes many of problems that arise from the arbitrariness of binning. As the name itself implies, the distance is based on the notion of the amount of effort required to convert one instrumental music to another based on the analogy of transporting mass from one distribution to another. If two instrumental music are viewed as distributions and view one distribution as a mass of earth in space and the other distribution as a hole (or set of holes) in the same space then the EMD [21] is the minimum amount of work involved in filling the holes with the earth. Some researchers analysis of the EMD argue that a measure based on the notion of a signature is better than one based on a histogram. A signature {sj = mj ,wmj } is a set of j clusters where mj is a vector describing the mode of cluster j and wmj is the fraction of features falling into that cluster.

Thus a signature is a generalization of the notion of a histogram where boundaries and the number of partitions are not set in advance; instead j should be 'appropriate' to the complexity of the instrumental music. The example in Figure 6.3 illustrates this idea. The clustering can be thought as a quantization of the instrumental music in some frequency space so that the instrumental music is represented by a set of cluster modes and their weights. In the figure the source instrumental music is represented in a 2D space as two points of weights 0.6 and 0.4; the target instrumental music is represented by three points with weights 0.5, 0.3 and 0.2. In this example the EMD is calculated to be the sum of the amounts moved (0.2, 0.2, 0.1 and 0.5) multiplied by the distances they are moved. Calculating the EMD involves discovering an assignment that minimizes this amount.

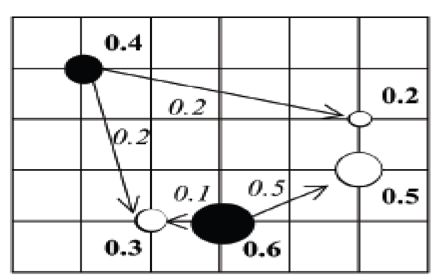

Figure 3: An example of the EMD between two 2D signatures with two points (clusters) in one signature and three in the other.

For two instrumental music described by signatures S = {mj ,wmj }nj =1 and Q = {pk,wpk}r k=1 . The work required to transfer from one to the other for a given flow pattern F:

$$WORK(S,Q,F) = \sum_{j=1}^{n}\sum_{k=1}^{r} d_{jk} f_{ik} \quad \text{---- (6)}$$

where djk is the distance between clusters mj and pk and fjk is the flow between mj and pk that minimizes overall cost. Once the transportation problem of identifying the flow that minimizes effort is solved by using dynamic programming. The EMD is defined as:

$$EMD(S,Q) = \frac{\sum_{j=1}^{n}\sum_{k=1}^{r} d_{jk}\ f_{jk}}{\sum_{j=1}^{n}\sum_{k}^{r} f_{jk}} \quad \text{-----(7)}$$

EMD is expensive to compute than linearly with the number of clusters. Nevertheless it is an effective measure for capturing similarity between instrumental music. It is identified that the EMD approach is giving better results than Cosine measure.

3.3 Model Experimentation

Any number data items may be considered for training. At present, 3 instrumental raagas for first experiment and 5 instrumental raagas for the second experiment are considered.

The classification in done using KNN in MATLAB with the following command

Class = knnclassify (Sample, Training, Group)

Sample= Matrix whose rows will be classified into groups. Sample must have the same number of columns as Training

Training= Matrix used to group the rows in the matrix Sample. Training must have the same number of columns as Sample. Each row of Training belongs to the group whose value is the corresponding entry of Group
Group= Vector whose distinct values define the grouping of the rows in Training

The classification of KNN assigns each row of Sample to the group for the closest row of Training. Group can be a numeric vector, a string array, or a cell array of strings. Training and Group must have the same number of rows. KNN classify treats empty strings in Group as missing values, and ignores the corresponding rows of Training. Class indicates which group each row of Sample has been assigned to, and is of the same type as Group.

EMD Experiment

CASE-1(Same Ordering of Raagas in training and sampling)

Sampling and Training with 3 Raagas

The given "sample" and "train" Command is applicable to sample and train the classifier. The following results are obtained on application of EMD [21] distance approaches. The given below figure classifies each row of the data in the given sample into one of the three groups in training.

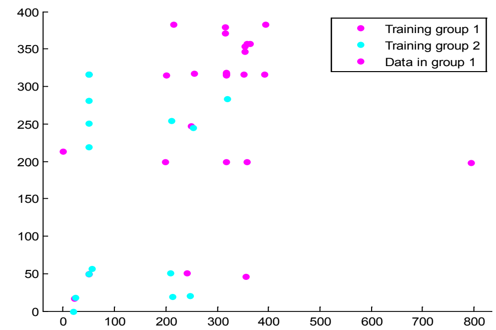

Figure 4: Scatter plot shows each row of the data in sample into one of the three groups in training.

CASE-2 (Different Ordering of Raagas in training and sampling)

The following figure utilize the same data as in Classifying Rows into One of Three Groups, but classifies the rows of sample using three nearest neighbours instead of one.

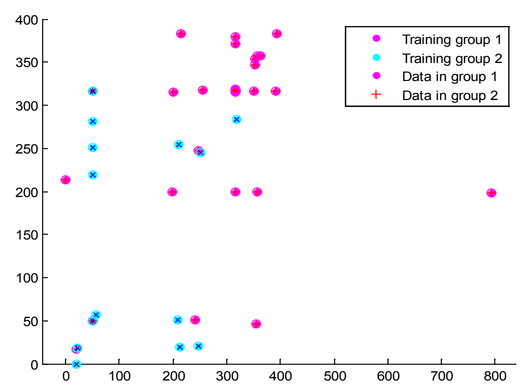

Figure 5: scatter plot shows the same data as in Classifying Rows into One of Three Groups.

If this plot is compared with the one in Classifying Rows into One of Three groups, you see that some of the data points are classified differently using three nearest neighbours.

Case-3(Different Raagas for training and sampling)

Sampling with 3 Raagas and Training with 5 Raagas

The given figure classifies each row of the data in sample into one of the five groups in training.

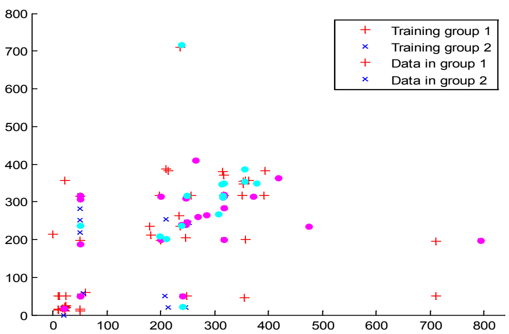

Figure 6: Scatter plot shows each row of the data in sample into one of the five groups in training.

Classifying Rows Using the Three Nearest Neighbors

From the above results, it uses the same data as in Classifying Rows into One of Two Groups, but classifies the rows of sample using three nearest neighbors instead of one. The system predicts the output which raaga it is or closer to training raaga.

COSINE Experiment

Sampling with 3 Raagas and Training with 5 Raagas

The following results are obtained by applying COSINE distance approach and the figure classifies each row of data in the sample into one of the five groups in training.

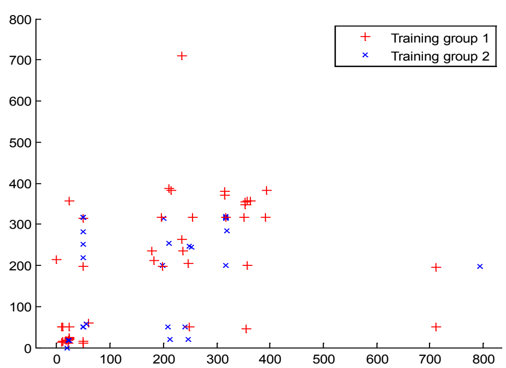

Figure 7: Scatter plot shows each row of the data in sample into one of the five groups in training.

Classifying Rows Using the Three Nearest Neighbors

The given example uses the same data as in Classifying Rows into One of Two Groups, but classifies the rows of sample using three nearest neighbors instead of one.

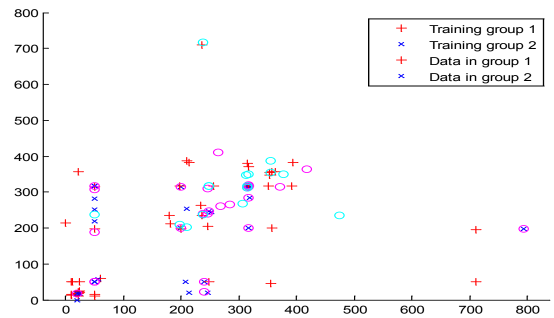

Figure 8: Scatter plot shows the same data as in Classifying Rows into One of Two Groups

Similarly, cosine approach is applied to all cases that were discussed in EMD and the comparison is shown in the results and also the discussion section.

4. RESULTS AND DISCUSSION

The input signal is sampled at 44.1 KHz. The identification of different Raagams for the purpose of evaluating this algorithm is considered. For the purpose of Raaga identification seven different instruments are considered. The signal is made to pass through the signal separation algorithm, and segmentation algorithm.

The result showing the segmentation points for one input is shown in the figures given. This is the first level of segmentation where the protruding lines indicate the points of segmentation. After identifying the segmentation points the frequency components are determined using the HPS algorithm and tabulated the frequency values which have the dominant energy. Using the raaga identification system, the confusion matrix is determined.

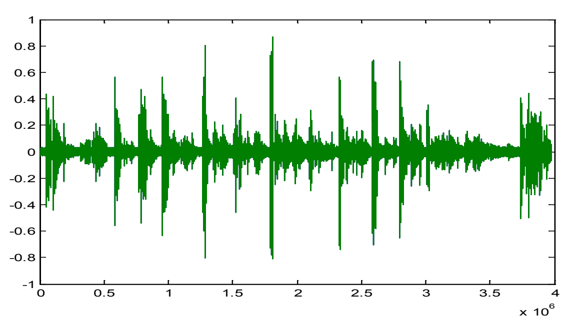

Figure 9: Plot Graph for Begada Raaga

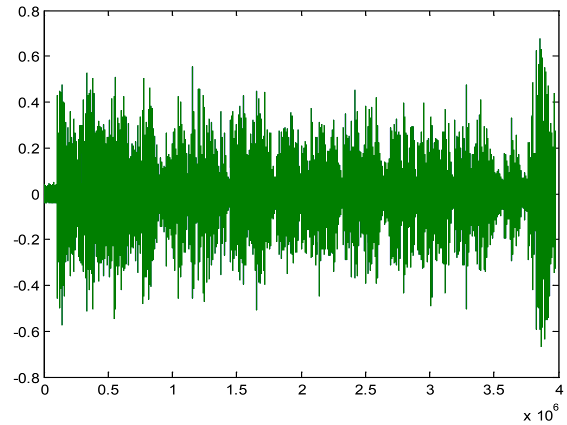

Figure 10: Plot Graph for Kharaharapriya Raaga

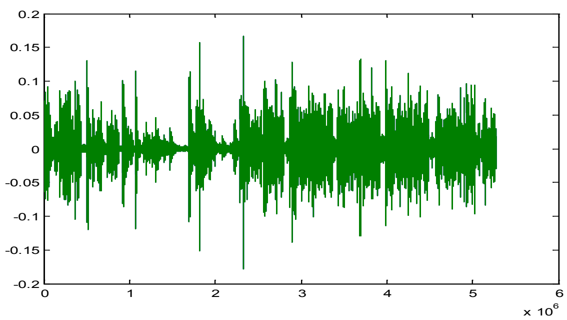

Figure 11: Plot graph for Aarabhi raaga

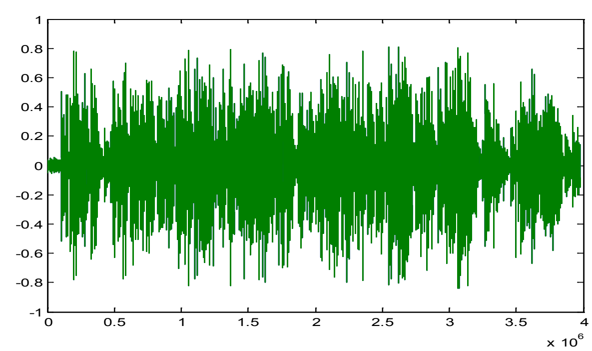

Figure 12: Plot file for Malahari Raaga

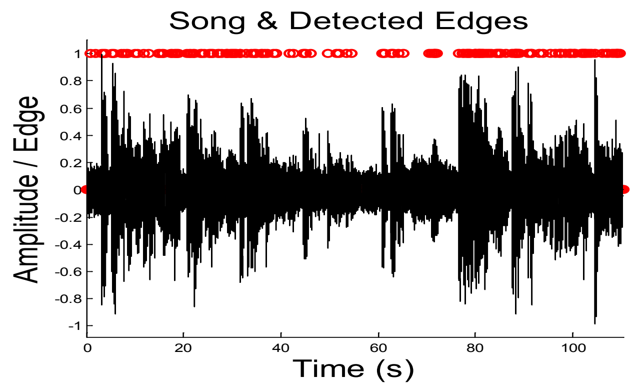

Figure 13: Bhiravi raaga for Edge detection

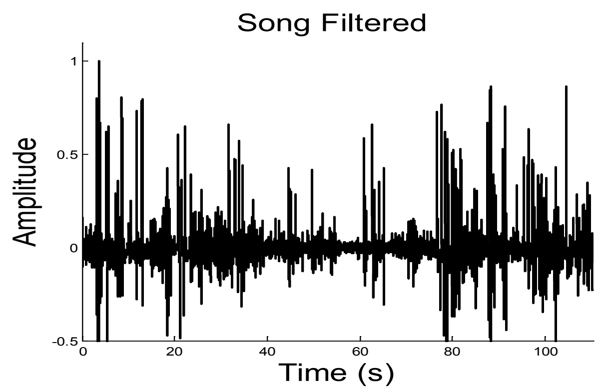

Figure 14: Bhiravi raaga for Song Filtere

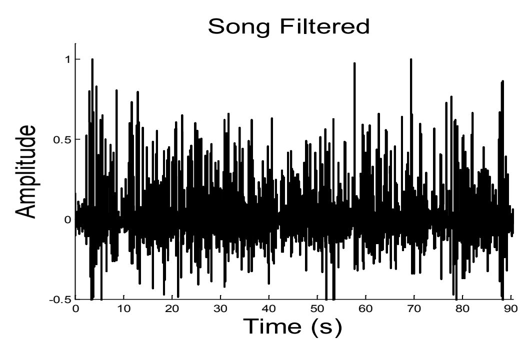

Figure 16: Malahari raaga for Song Filtered

The following are the results obtained by applying Cosine Distance measure

Cosine Distance: The Data is same for Train and Sample

Table 1: Confusion Matrix: Same data for Train and Sample

Cosine Distance: The Data is different for Train and Sample

Name of the Raaga	Recognized Raagas (%)					
	Beg ada	Vanas apathi	sunda vinodi ni	Desh	Hind ilom	Ravich andrik a
Begada	**94**	58	58	62	65	65
Vanasap athi	58	**94**	63	70	72	75
sundavi nodini	58	68	**94**	68	70	68
Desh	62	70	76	**94**	75	78
Hindilo m	65	72	70	85	**94**	80
Ravicha ndrik	65	75	68	78	80	**94**

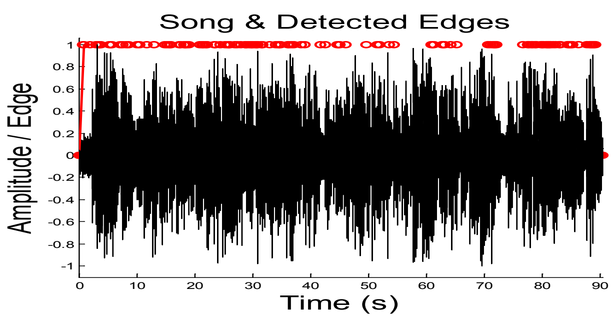

Figure 15: Malahari raaga for Edge detection

Table 2: Confusion Matrix: Different data for Train and Sample

Name of the Raaga	Recognized Raagas (%)					
	Beg ada	Vana sapat hi	sunda vinodi ni	Desh	Hindilo m	Ravichan drika
Sri	90	58	58	62	65	65
Bhirav i	58	88	63	70	72	75
Abheri	58	68	78	68	70	68
Malah ari	62	70	76	84	75	78
Sahana	65	72	70	85	86	80
Todi	65	75	68	78	80	88

The following are the results obtained by applying EMD Distance measure

EMD: The Data is same for both Train and Sample

Table 3: Confusion Matrix: Same data for Train and Sample

Name of the Raaga	Recognized Raagas (%)					
	Beg ada	Vana sapat hi	sund avin odini	Desh	Hindil om	Ravich andrik a
Begada	**98**	68	58	72	65	75
Vanasa pathi	78	**98**	63	80	82	65
sundavi nodini	68	78	**98**	88	70	78
Desh	72	70	76	**98**	85	88
Hindilo m	65	72	70	85	**98**	80
Ravicha ndrik	65	75	68	78	80	**98**

5. CONCLUSION

It is thus evident that K-NN is very simple to understand and easy to implement. It should therefore, be considered in seeking a solution to any classification problem. In some circumstances where, an explanation of the output of the classifier is useful, k-NN can definitely be very effective if an analysis of the neighbours is useful as explanation.

In order to improve classification process, an EMD approach is used for fast convergence. K-NN is very sensitive to irrelevant or redundant features because, all features contribute to the similarity and thus to the classification. This can be ameliorated by EMD approach and feature selection or feature weighted voting. The EMD results are compared to Cosine distance measure and observed that EMD apparently gives better results.EMD: The Data is different for Train and Sample

Table 4: Confusion Matrix: Different data for Train and Sample

Name of the Raaga	Recognized Raagas (%)					
	Begada	Vanasapathi	sundavinodini	Desh	Hindilom	Ravichandrika
Sahana	89	68	78	62	65	65
Todi	68	88	63	70	72	75
Sri	58	68	78	68	70	68
Bhiravi	72	70	76	84	75	78
Abheri	75	72	70	85	86	80
Malahari	70	75	68	78	80	88

6. REFERENCES

[1] Rajeswari Sridhar, Geetha T. V, “Swara Indentification for South Indian Classical Music”, ICIT '06 Proceedings of the 9th International Conference on Information Technology, IEEE Computer Society, ISBN:0-7695-2635-7.

[2] Youngmoo E. Kim, Brian Whitman” Singer Identification in Popular Music Recordings Using Voice Coding Features”, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.

[3] Paul G., Corine G.M., Christophe C., Vincent F. “automatic classification of environmental noise events by hidden Markov models”, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.52

[4] Berger. “Some factors in the recognition of timbre“. *J. Audio. Eng. Soc*. 30, pp. 396-406.

[5] Clark, Milner. “Dependence of timbre on the tonal loudness produced by musical instruments“. *J. Audio. Eng. Soc*. 12, pp. 28-31.

[6] Eagleson, H. W., Eagleson, O. W. “Identification of musical instruments when heard directly and over a public-address system“. *J. Acoust. Soc. Am*. 19, pp. 338-342.

[7] Strong, Clark. “Perturbations of synthetic orchestral wind instrument tones“. *J. Acoust. Soc. Am*., Vol. 41, pp. 277-285.

[8] Bhatkande.V (1934), Hindusthani Sangeet Paddhati.Sangeet Karyalaya, 1934.

[9] Schroeder.M.R (1968), ” Period histogram and product spectrum: New methods for fundamental-frequency measurement”, Journal of the Acoustical Society of America, , vol. 43, no. 4, 1968.

[10]. A. Prasad et al. “Gender Based Emotion Recognition System for Telugu Rural Dialects using Hidden Markov Models” Journal of Computing: An International Journal, Volume2, Number 6 June 2010 NY, USA, ISSN: 2151-9617

[11]. S. Dixon: “Multiphonic Note Identification”: Proc. 19th Australasian Computer Science Conference: Jan-Feb 2003.

[12]. W. Chai and B. Vercoe: “Folk Music Classification Using Hidden Markov Models”: Proc. Internation Conference on Artificial Intelligence: June 2001.

[13]. Tarakeswara Rao B et. All “A Novel Process for Melakartha Raaga Recognitionusing Hidden Markov Models (HMM)”, International Journal of Research and Reviews in Computer Science (IJRRCS), Volume 2, Number 2,April 2011, ISSN: 2079-2557.

[14]. A. Ghias, J. Logan, D. Chamberlin and B. C. Smith: “Query by Humming – Musical Information Retrieval in an Audio Database”: Proc. ACM Multimedia, pp. 231-236: 1995.

[15]. H. Deshpande, U. Nam and R. Singh: “MUGEC: Automatic Music Genre Classification”: Technical Report, Stanford University: June 2001.

[16] A. J. Viterbi: “Error bounds for convolutional codes and an asymptotically optimal decoding algorithm”: IEEE Transactions on Information Theory, Volume IT-13, pp.260-269: April 1967.

[17]. L. E. Baum: “An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes”: Inequalities, Volume 3, pp. 1-8: 1972.

[18]. L. E. Baum and T. Petrie: “Statistical inference for probabilistic functions of finite state Markov chains”: Ann.Math.Stat., Volume 37, pp. 1554-1563: 1966.

[19]. Gaurav Pandey, Chaitanya Mishra, and Paul Ipe “TANSEN: a system for automatic raga identification”

[20] Eagleson, H. W., Eagleson, O. W. "Identification of musical instruments when heard directly and over a public-address system". *J. Acoust. Soc. Am*. 19, pp. 338-342.

[21] Shingo Kuroiwa, Satoru Tsuge, Masahiko Kita, and Fuji Ren, [2007], "Speaker Identification Method Using Earth Mover's Distance for CCC Speaker Recognition Evaluation 2006", Computational Linguistics and Chinese Language Processing Vol. 12, No. 3, September 2007, pp. 239-254.