iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VI **Month of publication:** June 2019

DOI: http://doi.org/10.22214/ijraset.2019.6012

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

Marathi Spoken Word Recognition using Spectral Domain Features over Distributed Frequency Content

Vedant Shende[1], Chinmay Kunte[2], Sagar Darekar[3], Tushar Deshmukh[4]
[1, 2, 3, 4]*Computer Science and Engineering, Savitribai Phule Pune University*

Abstract: ***Word recognition is one of the important area in speech recognition. Local language spoken word recognition is the next step in the technological advancement. This paper presents the new approach for Marathi word recognition using Spectral Domain features like Mel-frequency cepstral coefficients (MFCC) and spectral centroid over distributed frequency content. A very little or no work is done in Marathi spoken word recognition. That is the reason why no standard Marathi word recognition tool was available online. A dataset of 40 native speakers from different region including 20 males and 20 females is developed for total 20 words related to grocery items. The selected words are uttered in Marathi language. Each speaker uttered a word 10 times. The creation of the dataset for the Marathi words was the main contribution of this work. To the best author knowledge, the SVM classifier is used for the first time to recognize Marathi words. Out of 40 speakers, 30 speakers feature set was used for training purposes and remaining of the 10 speakers feature set are used for testing. The results obtained from the purposed method are satisfactory and achieved 94.75 % overall accuracy.***

Keywords: Classifiers, Corpus development, Feature vector, Marathi language spoken words recognition, Spectral features.

I. INTRODUCTION

The world is moving towards automation, computerization and the interaction of human with machine is increasing day by day. The increasing interaction resulting in technological advancement in language recognition techniques. Further the local language recognition is getting importance in natural language recognition. Hence, the interaction of the human with machine must be natural and pervasive. In the recent years some work on local language recognition is reported. The languages Persian [1], Arabic [2], Hindi [3], Urdu [4], Spanish [5], Chinese [6], Tamil [7], Telugu [8] etc. are under focus in language recognition techniques. One of the important languages of India is Marathi with around 80-90 million speakers around the world. Marathi is widely spoken language in India. The state of Maharashtra in India has majority of Marathi speakers with Marathi as first language in the state. Therefore, the focus of the researcher is to develop robust and efficient model for Marathi words recognition. To the best of authors knowledge, very little work has been done on Marathi word recognition and the main reason behind this is the unavailability of standard corpus. No standard corpus of Marathi words is available online. Standard corpus plays a critical role in building, training and testing of proposed model. Initially the speech standard corpus was developed at MIT for English language known as TIMIT [9]. Similarly, some standard corpuses have been developed such as Japanese [16], French [17], Spanish [18], Arabic [19], Bengali [20], Tai [21] etc. Since no standard corpus for Marathi words recognition is available online, therefore the first task is to build a standard corpus for Marathi words. Hence, the database of Marathi words with 40 speakers including 20 males and 20 females aging from 16 to 70 years has been built. The proposed method is build, trained and tested on the corpus developed in this work. The results obtained from the proposed method are very satisfactory and achieved 94.75% of overall accuracy. Results obtained from different classifiers are discussed in Results and discussion section. Rest of the paper is organized as follow: Section 2 discusses related work. Experimental setup and development of standard corpus is covered in Section 3. Feature extraction and algorithm is discussed in Section 4. Classifiers are covered in Section 5. Results and discussion is done in section 6, followed by the conclusive remarks in section 7.

Fig. 1 Block diagram for the proposed system.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

II. RELATED WORK

A lot of research has been done on the speech recognition in last few decades. In early decades Bells lab researchers investigated the science of speech perception [10]. The three Bells lab researchers build a system for single digit recognition system. In [11], the authors wrote a research letter on speech recognition. Speaking process and understanding the fluency of the speech was discussed. A special system was built to recognize four (4) vowels and nine (9) consonants. The recognition of vowels and consonants highlighted the first utilization of statistical syntax in automatic speech apperception to achieve such abilities in local languages is the first priority of advance technology so that communication with computers is easy for every individual in a natural way. A paper was presented on digit recognition system in Persian language in which hidden Markov model is used to decompose the words into small parts for solving the problem arising from the pronunciation [12]. The first positive result of word recognition came into existence in 1970 when the general pattern techniques were introduced by Kumar et al [13]. Some research work is also available on the system other than this sphinx based system. Pattern matching of Urdu speech recognition using acoustic Phonetic Modelling approaches is also discussed. Similarly, research was done on Hindi language and the overall accuracy achieved was 87.01%. In [14] the practical direction was given on the development of robust acoustic models utilizing the difference tools like Hidden Markov Model Toolkit and Sphinx tools. In this paper the speech apperception problem is analysed and developed a system which uses the Mel frequency Cepstral coefficients (MFCC) technique [15]. Speaker identification is also a separate field of research from the last few years. There are a lot of problems in that field.

III.EXPERIMENTAL SETUP

A. *Corpus Development*

Since no standard corpus is available online so the first stepSince no standard corpus is available online so the first step is to develop the Marathi word corpus to build, train and test the proposed model. Following steps are taken to develop a standard corpus.

B. Recording Speeches

All the recordings are done on mobile handset in .wav format. The recordings are done in a noise free environment. All the recordings are done from the native speakers from different regions of Maharashtra to get different accents and most of them people of age group 15 to 65 years. Each speaker is asked to utter the Marathi words with some pause, then audacity software is used for editing and splitting.

C. Speaker Statistics

In order to develop a standard corpus, 40 native speakers were selected. For diversity 20 males and 20 females of different ages ranging from 15 to 65 years were included. Table 1 shows speaker classification for the proposed system.

TABLE I Speaker Statistics

Number of speakers	Male	Female
40	20	20

IV. FEATURE EXTRACTION AND ALGORITHM

The spectral features (frequency based features), which are obtained by converting the time based signal into the frequency domain using the Fourier Transform, like: fundamental frequency, frequency components, spectral centroid, spectral flux, spectral density, spectral roll-off, etc. These features can be used to identify the notes, pitch, rhythm, and melody.

For speech recognition, MFCC is widely used because of its accuracy and efficiency. In this work, the MFCC is used along with spectral centroid for spectral feature extraction and Marathi word recognition. The spectral centroid indicates where the" center of mass" of the spectrum is located.

A. Find Start and End of Required Audio

The audio file is in .wav format may contain noise. The segment of the .wav file which contain the part of uttered word will be required to get the optimized output. For that the start and the end of that segment needs to be calculated. Following formulas are used in python to get the start and end of the required wave segment.

1) Start trim = plotwave.detect leading silence (sound,dbel,50)
2) End trim = plotwave.detect leading silence (sound,dbel,50)
3) End trim = duration end trim
4) New start = int ((start trim * signal length) / duration)
5) New end = int ((end trim * signal length) / duration)

B. Normalize the Wave

Each recorded file of a word may be having different amplitude. The amplitude of the wave should be normalized between -1 to 1. To normalize the amplitude, the wave should be divided by maximum amplitude in the wave. Following formulas are used to normalize the audio file.

1.signal = signal / max(signal) # Gives maximum amplitude in python

C. Find Feature Values for MFCC

1) *Preemphasis*: In order to smoothen the spectrum, the preemphasis is used. The aim is to boost up the high frequencies that are suppressed in human sound production mechanism.

Preemphasis is done using the equation given below:

$Y[n] = X[n] - \alpha X[n]$

where, Y [n] is the output signal with the boosted frequencies.

X[n] is the input Marathi grocery words and in this work is selected as 0.97.

2) *Framing:* Since, speech is a non-stationary signal and to make it stationary, the speech signal is divided into small segments, which is known as Framing. The Faming is achieved by multiplying a time window function with the speech signal. The length of the window is kept 1/10 of total signal length in such the speech is almost stationary with 0% overlapping. Following functions are used to calculate the mfcc values for the given wave in python.

a) For signal in divided signal:

b) Temp values = numpy.mean(mfcc(y=signal, sr=16000, n mfcc=12).T, axis=0)

c) mfcc values = numpy.append(temp values[2:], mfcc values)

Where, divided signal is array of segments containing each frame of signal.

sr is the sampling rate.

n mfcc is number of mfcc elements per frame.

Above mentioned method gives 12 mfcc values per frame. The proposed algorithm has total 10 frames per audio segment. Hence, there are total 120 mfcc feature values.

D. Find Feature values for Spectral Centroid

Spectral centroid for the .wav file shows where the mass of the wave is concentrated. For the proposed system the center of mass is calculated for each segment or frame. The actual audio segment in the wave file is divided into 10 segments and spectral centroid is calculated for each segment. Following is the implementation for the spectral centroid in python.

1) Def spectral centroid(x, samplerate=44100):

2) Magnitudes = np.abs(np.fft.rfft(x)) # magnitudes of positive frequencies

3) Length = len(x) freqs = np.abs(np.fft.fftfreq(length, 1.0/samplerate)[: length//2+1]) # positive frequencies

4) Return np.sum(magnitudes*freqs) / np.sum(magnitudes)

E. Make Feature Vector for Training and Testing

After the implementation of mfcc and centroid, 120 values of mfcc and 10 values of spectral centroid will be obtained for a Marathi word. These 130 values will act as a feature vector for one word. These set of values will be given as input to the neural network. In the proposed system, there will be total 120 feature vectors for 20 Marathi words selected. So, the training matrix will have 20 rows and 131 columns the last column will contain the output for the feature vector. In the corpus development there are total 40 speakers to utter a word 10 times. Out of the total data 75% of data is given as input for training and 25% is given for testing for different classifiers.

F. Algorithm for Speech Recognition

Input: AudioSignal

Output: Textdata

1) Start Procedure

2) Accept audio input

3) Remove noise

4) Normalize wave

5) divide wave into 10 segments
6) Find spectral feature values [mfcc, centroid]
7) Use feature value for training
8) Use the trained model for word recognition
9) End procedure

V. CLASSIFIERS

The classification in the proposed method is in fact the matching of features extracted from the test Marathi words and the features saved in the database in training phase. In this work two classifiers are studied i.e. SVM and KNN. In machine learning, support vector machines are widely used for feature matching and classification. The principle of SVM is to maximize the functional margin between nearest training data of distinct class and construct an optimal hyper plane. KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970s as a non-parametric technique. The distance function used in this paper is Euclidean distance.

VI. RESULTS AND DISCUSSION

The paper focused on the implementation of Marathi word recognition system 20 words related to grocery shop. First the databases of 40 Marathi native speakers were built, followed by feature extraction. For dimensionality reduction, statistical feature selection technique is used to reduce the feature set. In last step the classification is performed by using two different classifiers i.e. SVM and KNN. For classification purposes the feature set is divided into two sets i.e. training and testing. 75% of the data (30 speakers) were used for training and remaining 25% of the data (10 speakers) were used for testing the system. Extraction of features and classification (KNN and SVM) were implemented in Python environment. The accuracy achieved by Marathi word recognition with 20 speakers' database having 75% of training and 25% of testing data is 83.8% with KNN classifiers. When the dataset is increased to 40 speakers with 75% training and 25% testing the accuracy of Marathi word recognition raised to 91.50% with KNN classifier. The accuracy is further improved to 94.75% when the classifier is replaced with SVM. The confusion matrix with different classifiers and dataset are given in figure 2-3. Table 2 contains the summaries of results obtained by different classifiers.

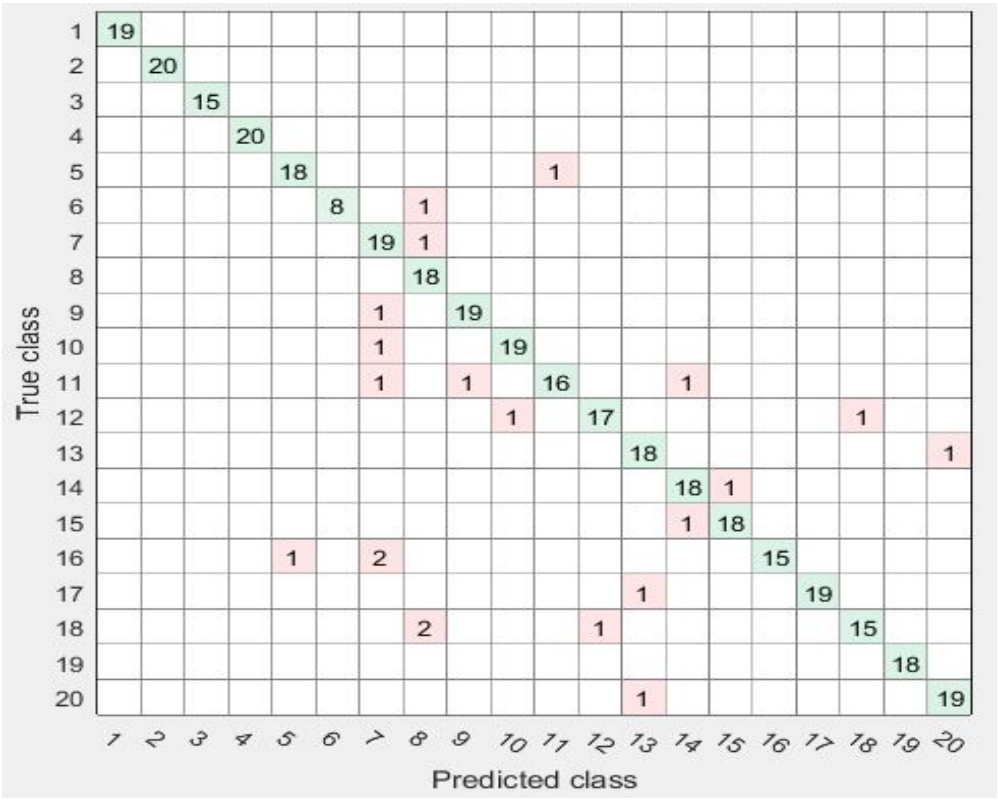


Fig 2. Confusion matrix for SVM classifier Accuracy 94.75%

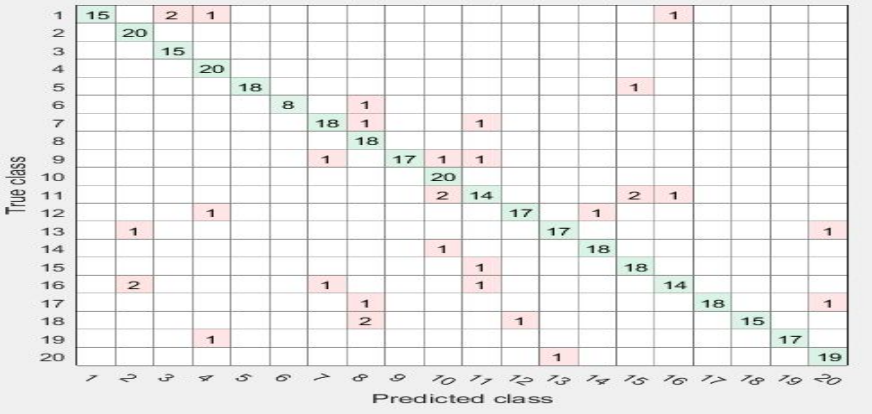


Fig 3. Confusion matrix for KNN classifier Accuracy 91.50%

TABLE III
Summary of the results of proposed system

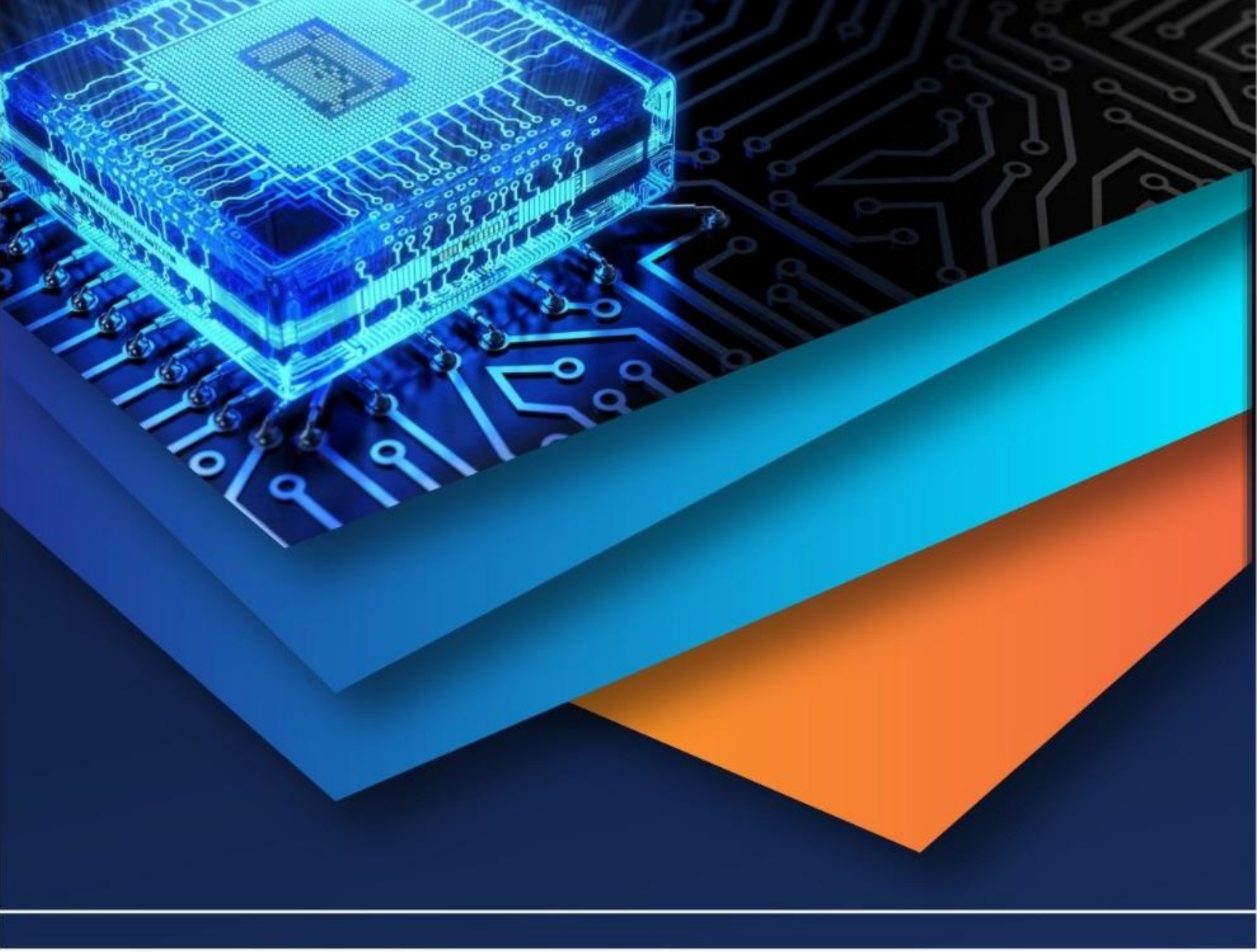
Feature Extraction	Feature Selection	Classifier	Recognition Rate(%)
Spectral Features	MFCC, Spectral Centroid	KNN	91.50
Spectral Features	MFCC, Spectral Centroid	SVM	94.75

VII. CONCLUSION

In this paper Marathi word recognition using spectral feature extraction with different classifiers have been studied. Since no standard corpus for Marathi words are available online so the first task was to build a standard corpus. One of the main contributions of this paper is the creation of corpus. In feature extraction module, different spectral features have been extracted to train and test the classifier. To the best of authors knowledge, SVM is used for the first time for Marathi words classification and compared with KNN classifier. The results obtained from the proposed method are satisfactory and achieved an overall of 94.5% accuracy. In future, the authors are planning to extract some more unique features in order to improve the accuracy further while keeping the efficiency intact.

REFERENCES

[1] H. Liu, H. Motoda, and L. Yu, Feature selection with selective sampling, in Nineteenth International Conference Machine Learning, pp. 395402,2002.
[2] S. Das, Filters, wrappers and a boosting-based hybrid for feature selection, in International Conference on Machine Learning, vol. 1, pp. 7481, 2001.
[3] R. Kohavi and G. H. John, Wrappers for feature subset selection, Art. Intell., vol. 97, no. 1, pp. 273324, 1997.
[4] J. R. Quinlan, C4. 5: programs for machine learning. Elsevier, 2014.
[5] A.Ng, Cs229 lecture notes, 2012. [Online]. Available: http://cs229.stanford.edu/notes/cs229-notes5.pdf
[6] M. A. Hall, Correlation-based feature selection for machine learning, Ph. D. dissertation, The University of Waikato, 1999.
[7] J. Lin, Divergence measures based on the shannon entropy, IEEE Trans. Inform. Theory, vol. 37, no. 1, pp. 145151, 1991.
[8] M. Dash and H. Liu, Consistency-based search in feature selection, Art. Intell., vol. 151, no. 1, pp. 155176, 2003.
[9] E. Alpaydin, Introduction to machine learning. MIT press, 2014.
[10] M. A. Hall, Correlation-based feature selection of discrete and numeric class machine learning, Proc. 17th Intl Conf. Machine Learning, pp.359-366, 2000.
[11] M. Dash, H. Liu, and H. Motoda, Consistency based feature selection, in Pacific-Asia conference on knowledge discovery and data mining, pp.98109, 2000.
[12] T. Howley, M. G. Madden, M.-L. Oonnell, and A. G. Ryder, The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data, Knowledge-Based Systems, vol. 19, no.5, pp. 363370, 2006.
[13] S. Hettich and S. D. Bay, The uci kdd archive, 1999. [Online]. Available: http://kdd.ics.uci.edu
[14] J. Kacur and G. Rozinaj, Practical issues of building robust HMM models using HTK and SPHINX systems. INTECH Open Access Publisher, 2008.
[15] L. Muda, M. Begam, and I. Elamvazuthi, Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques, arXiv preprint arXiv:1003.4083, 2010
[16] W. B. Powell, Approximate dynamic programming: solving the curses of dimensionality. John Wiley Sons, 2007, vol. 703.
[17] R. Kohavi and G. H. John, Wrappers for feature subset selection, Art. Intell., vol. 97, no. 1, pp. 273324, 1997.
[18] D. W. Aha, Tolerating noisy, irrelevant and novel attributes in instancebased learning algorithms, Int. J. Man-Machine Studies, vol. 36, no. 2, pp. 267287, 1992.
[19] P. Clark and R. Boswell, Practical machine learning tools and techniques with java implementation. Morgan Kaufmann Publishers, 2000.
[20] P. Langley, Selection of relevant features in machine learning, in Proceedings of the AAAI Fall symposium on relevance, vol. 184, pp.245271, 1994.
[21] K. Kira and L. A. Rendell, The feature selection problem: Traditional methods and a new algorithm, in National Conference on Artificial Intelligence San Jose, vol. 2, pp. 129134, 1992.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY