Phonotactic Model for Spoken Language Identification in Indian Language Perspective

Sanghamitra Mohanty
Department of Computer Science and Application
Utkal University, Bhubaneswar, Odisha, India -751004

ABSTRACT

Indian Languages are Indo-Aryan being influenced by Sanskrit or Dravidian being influenced by Tamil. Dravidian Languages have the influence of Sanskrit also. All Indian Languages have the influence of Pali language for which the graphemes are being influenced Brahmi. All the Indian languages are phonetic in nature. Every Indian language has its distinctive phone sets. North Indian languages are Indo- Aryan and South Indian Languages are Dravidian. Considering their respective Phonetic properties during speaking we have tried to consider the special CV behaviour of the language in their syllables and are able to identify the Language analysing it with the limited training data set available using the SVM Classifier. During this process we have analysed the PPR Language Modelling concept for four major Indian languages like Hindi, Bengali, Oriya, and Telugu and the results are quite appreciable.

General Terms

Spoken Language Identification, Speech Processing, Support Vector Machine

Keywords

LID, Indian Language, Support Vector Machine, Phonotactic

1. INTRODUCTION

Language Identification (LID)[1-4] has drawn the attention of speech scientists in the last one decade as it is one of the needs in the process of Multilingual Continuous Speech Recognition and Speech – To –Speech Translation. It involves the task of identifying the language from a short duration of speech signal uttered by any speaker. Different approaches are made for the LID experimentation. A very few attempts are made for Indian Languages[5,6]. India is a multilingual and multicultural country. Indian Languages are having their special behaviour as all Indian languages are phonetic in nature. Two types of languages are spoken in India. One is of Indo-Aryan family and the other is the Dravidian family. Indo-Aryan family languages like Hindi, Bengali, Oriya, Gujurati, Punjabi and Marathi etc. come under Indo-Aryan family while Tamil, Telegu, Malayalam and Kannada are under the Dravidian family. Indi-Aryan Languages have the influence of Sanskrit while Dravidian languages are having the influence of Tamil and Sanskrit. In the Grapheme from both the families are having the influence of Brahmi scripts during writing. These Languages have more number of dialects, but we have considered the main form of the four languages considered. Some people of the North eastern part of India also speak Indo-European, Mon-Khmer, and Sino-Tibetan. In our experiment we have considered five Languages with an attempt to consider the influence on the major languages with distinct scripts. Those are Hindi, Bengali, Oriya among Indo-Aryan and Telugu among the Dravidian languages.

In the process of recognition with respect to Spoken Languages it is observed that human are the best identifier of a language. However during automatic recognition we have to consider several factors like with respect to language identification **o**ne language differs from another language in one or more of the following: [7-12]

(a) Phonology : Phone sets would be different for each of the languages.

(b) Morphology : The word roots and the lexicons may be different for different category of languages.

(c) Syntax : The sentence patterns with respect to grammar are different.

(d) Prosody : Duration, pitch, and stress patterns vary from language to language.

In this paper attempt has been made to identify a language while analysing the language at the CV segment level. The specific units are analysed using the Parallel Phone Recognition Language Model (PPRLM) using the statistical classifier SVM with Radial Basis Function (RDF) in a limited training set. In Section two we have analysed the phonetic behaviour of all the languages considered here. Section three has the discussion on the SVM classifier while Section four has the Experiment and Results. In Section five the Error Matrix for all the five languages with respect to phone level, syllable level and finally word level are presented. Section Six has the conclusion.

2. CONSONANTS AND VOWELS OF INDIAN LANGUAGES

Indian languages are phonetic in nature. The grapheme to phoneme mapping is linear. All the languages have their own set of Vowels and consonants. They have a good mapping from one language to the other with respect to the features like pitch or fundamental frequency, duration of utterance of a standard CV unit and moreover the absence of some of the consonants in the typical Dravidian language Tamil. Annexure – I describes the Phone i,e, the Vowels and Consonants of four of the languages namely Hindi, Odiya, Bengali and Telegu.

3. PHONETIC BEHAVIOUR OF INDIAN LANGUAGES

While making a study on the phonetic behaviour of the languages we have to make an analysis of the features of the language specifically with respect to its utterance style. Considering it we have observed that the features are mainly discussed below. During the analysis of the features of any language it is important to note that the linguistic distinctive features is the building blocks of phones in the sense that any phone can be defined by its distinctive features. These features can be coded as present or absent. These features can also be categorised with respect to its manner and place of utterance. In this paper we are concerned with one manner feature and ten place features [13]. The manner feature is the sonorant feature, which is essentially a measure of resonance in the vocal tract. And Sonorant phones are generally the vowels, glides, and nasals.

Next are the place features which are the articulatory bound and depend upon the location of the physical articulators in the vocal tract. From among the articulatory we consider the following features while considering their articulation:

Alveolar fricative: Sounds made by air flowing around the tongue when the tongue tip is pressed against the alveolar ridge /s/.

Alveolar nasal: Sounds made by opening the velo-pharyngeal port when the tongue tip is pressed against the alveolar ridge /n/.

Alveolar stop: Sounds made by the release of a closure in the vocal tract when the tongue tip is pressed against the alveolar ridge /d/.

Inter-dental fricative: Sounds made by air flowing around the tongue when the tounge body is between the teeth /th/.

Labial nasal: Sounds made by opening the velo-pharyngeal port when the lips are pressed together /m/.

Labial stop: Sounds made by the release of a closure in the vocal tract when the lips are pressed together /b/.

Labio-dental fricative: Sounds made by air flowing around the top teeth when they are pressed against the lips /f/.

Post-alveolar: Sounds made by pressing the tongue body against the palate behind the alveolar ridge /ch/.

Retroflex: Sounds made by creating a resonant cavity in the mouth below the tongue /r/. Velar Stop: Sounds made by the closure and release of the glottis /k/.

4. CORPORA CREATION

We have created the corpora for these four languages the voice of using 600 phrases of day to day talking consisting of 10 words in average per phrase. 25 males and 15 females within the age range of 23 to 48 have given their voice for the text fragments. The recording is done in the laboratory environment using noise cancellation microphone. The sampling rate is 16 bit in single channel of 16000 Hz.

5. SVM CLASSIFIER

Support Vector Machine (SVM) [14-18] is a stable classifier developed so far. It can be used for any type of statistical data handling with probabilistic values or a pattern matching from a training set. The classification is done using the dynamic separator. During classification we can have many linear separators. But it becomes difficult which separator to use as the difference is marginal from one set to the other set of values. In SVM we can have a distance from input **x***i* to the separator and the one that is

$$\mathbf{r = (w^T x_i + b)/|w|} \quad (1)$$

closest to the hyperplane are the support vectors. Margin ρ of the separator is the distance between support vectors. Thus during classification a group will have a range instead of a crisp unit. Mathematically we can have the description of the SVM as given below.

(i) Let training set $\{(\mathbf{x}i, yi)\}i=1..n$, $\mathbf{x}i\in\mathbf{R}d$, $yi \in \{-1, 1\}$ be separated by a hyperplane with margin ρ. Then for each training example ($\mathbf{x}i$, yi):

$$\boldsymbol{yi(\mathrm{wTx}i + b) \geq \rho/2} \quad \mathbf{(2)}$$

(ii) For every support vector **x***s* the above inequality is an equality. After rescaling **w** and *b* by *ρ*/2 in the equality, we obtain that distance between each **x***s* and the hyperplane is

$$\mathbf{r = y_s(w^T xs + b)/|w| \ = \ 1/|w|} \quad \mathbf{(3)}$$

(iii) Then the margin can be expressed through (rescaled) **w** and b as:

$$\mathbf{\rho=2r=2/|w|} \quad \mathbf{(4)}$$

This is an effective prior for avoiding over-fitting, which results in a sparse model dependent only on a subset of kernel functions. The extension to non-linear boundaries is acquired through the use of kernels that satisfy Mercer’s condition. The kernels map the original input vector *x* into a high dimension space of features and then compute a linear separating surface in this new feature space. In practice, the mapping is achieved by replacing the value of dot production between two data points in input space with the value that results when the same dot product is carried out in the feature space [14-18]. The following is formations

$$\mathbf{Max(\sum \alpha_i + \sum \alpha_i \alpha_j \; y_i y_i K(x_i,x_j))} \qquad (5)$$

$\boldsymbol{\alpha \quad i \qquad i,j}$

The kernel function K defines the type of decisio surface that the machines will build. In our experiments, the radial basis function (RBF) kernel is used and it takes the form:

$$\boldsymbol{k(x_i,x_j) = \exp[(-1/2)(|x_i-x_j|/\sigma)^2]} \qquad (6)$$

Where σ is the width of the radial basis function.

6. EXPERIMENT AND RESULT

We have done the experiment with five Indian languages namely Hindi, Bengali, Oriya and Telegu. These languages are selected as the phone sets are more or less same for the first four languages, where Telegu has some more fricatives in its phone sets.

For the analysis a corpora consisting of texts of these four languages are developed. For it 5 speakers for each language are selected. Each one of them read a text on their language consisting of approximately 500 words. So the corpora have limited size of 2000 words with having the speciality of their own language with respect to vowels and consonants for the CV extraction [19].

The CV units are selected for the duration of 0.15 msecs [19]. As Hindi has a short vowel at the end of a word and Bengali has *swa* while Oriya uses a full unit of vowel it becomes easier for the classification during the input feeding for the SVM training. Telegu has the full CV units but with pitch higher than the other languages as it is more vocalic and aspirated in nature. So the pitch has significant distinctive characteristics.

We have tried to implement the PPR system [20], which is designed to recognize the four languages. It has a phonotactic[21,22] method based on acoustic models, which acts as the phone recognizer and secondly the set of language classifier model. A basic PPR system, designed to recognize any number of languages, contains two types of models: firstly a set of acoustic models that function as phone recognizers, and secondly a set of language models or classifiers that characterize the observed phonotactics of each language when recognized in terms of each of the phone recognizers. Here the CV unit of each language plays an important role for the language unit concerned. For phone matching we apply the Viterbi algorithm as given below.

Results of the experiments are presented and the confusion matrix in (%) are noted for CV level in Table - 1, in Vowel level in Table – 2 and in Word level in Table -3.

Table- 1 Confusion Matrix in (%) using SVM at CV level with Average recognition is 86.5%.

Language class	Oriya	Hindi	Bengali	Telegu
Oriya	93	6	8	7
Hindi	5	90	9	2
Bengali	9	7	75	2
Telegu	6	2	3	88

Table- 2 Confusion Matrix in (%) using SVM at the Vowel level with Average recognition is 81.5%.

	Oriya	Hindi	Bengali	Telegu
Oriya	85	6	8	7
Hindi	5	82	9	2
Bengali	9	7	75	2
Telegu	6	2	3	84

Table – 3 Confusion matrix in (%) using SVM at word level. Average recognition is 89.75%.

Language class	Oriya	Hindi	Bengali	Telegu
Oriya	93	8	9	3
Hindi	6	92	7	2
`Bengali	3	7	84	2
Telegu	2	2	3	90

7. CONCLUSION

It is observed from the experiment performed using the phonotactic principle taking into consideration the phonetic behaviour like manner and place of utterance of the languages and SVM as classifier are performing well and thus can be accepted. The confusion matrix is also drawn with respect to the CV combination for all the four languages Oriya, Hindi, Bengali and Telegu. The languages Oriya, Hindi and Telegu match well with the CVs, the vowels and also the word level and the results are noted in the tables above. It is observed that language identification can be done using SVM technique. For Bengali language the result is not that satisfactory as it has *swa* and needs special care to be taken before the CV matching are done making *swa* deletion.

Work is in progress for South Indian Languages in the next phase giving special emphasis to Tamil as many consonants are not uttered in this languages unlike Hindi, Telegu or Oriya. This may help to identify the South Indian language class as a special category from which the language identification (LID) will lead to a successful approach in Indian Language Perspective.

8. REFERENCES

[1] X. Huang, et al, “Spoken Language Processing”, Prentice Hall PTR, NJ, 2001.

[2] Jelinek. F, “Statistical Methods for Speech Recognition”, MIT Press, Cambridge, 1997.

[3] Rabiner, L.R, Schafer, R.W, “Digital Processing of Speech Signals”, Pearson education, 1st Edition, 2004.

[4] O’Shaughnessy, D, “Speech Communications Human and Machine”, Universities Press, 2nd Edition, 2001.

[5] Mohanty, S. and Swain , B. K. “Language Identification using Support Vector Machine”, Proceedings of O-COCOSDA-2010, Nepal, 2010.

[6] Mohanty, S., Bhattacharya, S., Bose, S., Swain, S., “An Approach To Parametric based Mood Analysis In Oriya Speech Processing” ,Proceedings of the International Symposium Frontiers of Research on Speech and Music(FRSM-2005).

[7] M.A. Zissman, ”Comparison of Four Approaches to Automatic Language Identification of Telephone speech, IEEE Transactions on Speech and Audio Processing”,1996.

[8] Navratil. J, ”Spoken Language Recognition - A Step Toward Multilinguality in Speech Processing”, IEEE Transactions on Speech and Ausio Processing, Sept. 2001.

[9] Muthusamy, Y.K, et al, ”Reviewing Automatic Language Identification”, IEEE Signal Processing Magazine, 1994.

[10] Schultz.T, et al, ”Language Independent and Language Adaptive Large Vocabulary Speech Recognition”, Proc. EuroSpeech, 1999, Hungary.

[11] Schultz, T and Kirchhoff, K “Multilingual Speech Processing”, Academic Press, 2006.

[12] Mak. B, et al, “Multilingual Speech Recognition with Language Identification”, Proc. ICSLP 2002.

[13] Ken Stevens, “Acoustic Phonetics”, MIT Press, Cambridge, MA, 1999.

[14] V. Vapnik. “The Nature of Statistical Learning Theory”. Springer-Verlag,1995.

[15] R. Duda, P. Hart, and D.Stork, “*Pattern Classification”*, Wiley, New York, 2001.

[16] N. Smith, M. Niranjan, “Data-dependent kernels in SVM classification of speech patterns”, in: Proceedings of the International Conference on Spoken Language Processing (ICSLP), Vol. 1, Beijing, China, 2000.

[17] William M. Campbell, Joseph P. Campbell, Douglas A. Reynolds, E. Singer, and P. A. Torres-Carrasquillo, “Support vector machines for speaker and language recognition” *Computer Speech and Language*, vol. 20, no. 2-3, 2006.

[18] OSU-SVM website: http://svm.sourceforge.net/license.shtml

[19] Praat software website: http://www.fon.hum.uva.nl/praat/.

[20] A. Montero-Asenjo, D.T. Toledano, J. Gonzalez-Dominguez, J. Gonzalez-Rodriguez, and J. Ortega- Garcia, “Exploring PPRLM performance for NIST 2005 language recognition evaluation,” in *IEEE Odyssey 2006:The Speaker and Language Recognition Workshop*, 2006.

[21] Keshet,J., Bengio, S. “Automatic Speech and Speaker Recognition Large Margin and Kernel Methods”, John Wiley and Sons, Ltd, Publication,1st edition, 2009.

[22] Pavel Matejka, Petr Schwarz, Jan Cernock, and Pavel Chytil, “Phonotactic language identification using high quality phoneme recognition,” in *Interspeech*, 2005.

Annexure – I

HINDI	**ORIYA**	**BENGALI**	**TELUGU**
VOWEL			
अ	ଅ	অ	అ
आ	ଆ	আ	ఆ
इ	ଇ	ই	ఇ
ई	ଈ	ঈ	ఈ
उ	ଉ	উ	ఉ
ऊ	ଊ	ঊ	ఊ
ऋ	ଋ	ঋ	ఋ
ऍ			
ऎ			ఎ
ए	ଏ	এ	ఏ
ऐ	ଐ	ঐ	ఐ
ऑ			
ऒ			ఒ
ओ	ଓ	ও	ఓ
औ	ଔ	ঔ	ఔ

CONSONANTS

क	କ	ক	క
ख	ଖ	খ	ఖ
ग	ଗ	গ	గ
घ	ଘ	ঘ	ఘ
ङ	ଙ	ঙ	ఙ
च	ଚ	চ	చ
छ	ଛ	ছ	ఛ
ज	ଜ	জ	జ
झ	ଝ	ঝ	ఝ
ञ	ଞ	ঞ	ఞ
ट	ଟ	ট	ట
ठ	ଠ	ঠ	ఠ
ड	ଡ	ড	డ
ढ	ଢ	ঢ	ఢ
ण	ଣ	ণ	ణ
त	ତ	ত	త
थ	ଥ	থ	థ
द	ଦ	দ	ద
ध	ଧ	ধ	ధ

न	ନ	ন	న
ऩ			
प	ପ	প	ప
फ	ଫ	ফ	ఫ
ब	ବ	ব	బ
भ	ଭ	ভ	భ
म	ମ	ম	మ
य	ଯ	য	య
र	ର	র	ర
ऱ			ఱ
ल	ଲ	ল	ల
ळ	ଳ		ళ
ऴ			
व			వ
श	ଶ	শ	శ
ष	ଷ	ষ	ష
स	ସ	স	స
ह	ହ	হ	హ