IJCSN International Journal of Computer Science and Network, May 2014
ISSN (Online) : 2277-5420
www.IJCSN.org
Conference Proceeding – RDFCORDGE - 2014

Automatic Language Identification Using Speech Processing: A Review

[1]Khagesh Kumar Sahu, [2]Mr. Vinay Jain

[1]M. Tech. Scholar, SSCET Bhilai.

[2]Associate professor, E&TC, SSCET Bhilai.

Abstract- Identifying the spoken language automatically is one of the area of research which attracted a lot of researchers. In the last few years various different algorithm have been proposed by different researchers. This paper present a brief discussion on various methods proposed by researchers for implementing language identification.

Keywords- *Multilingual speech recognition, Language identification, Phonology, Vector quantization, Autocorrelation.*

1. Introduction

Language identification is one of the most important speech processing operation which is used in various speech based application such as translation of spoken language, Multilangual speech recognition[1], retrieval of spoken document[2].In the past alot of research work has been done in this field and in the last decades this field witness a significant progress. For a human being, it is easy to identify the spoken language if the human being is familiar with the spoken language. But for the machines or computer, it is very difficult to identify the spoken language. Language identification system takes the speech of any language as input and with the help of some features of the speech signal, identify the language. Morphology, Prosody, phonology and syntax are some of the characteristics that make one language differ from other.

2. Past Approaches and Algorithms

A brief description of major algorithm and approaches used in language identification method are described here.
The first attempt in identifying the different language based on spoken speech features was done by Texas Instrument which was based was computing and identifying the frequency occurrences of some reference sounds which are different in different language. Once these reference sounds are Separated [3] from a speech signal then different language can be identified. The result of this method was very satisfying and encouraging with 64% correction percent for the test set carried out for seven different languages. Later on human interactive approach was also used to detect or determine the correct reference sound and hence 80% accuracy was achieved in paper[4]. The main drawback of determining the reference sounds manually was that it degrades the performance of the system with the addition of more language[4].This was depicted in [4] where inclusion of seven language deteriorated the accuracy from 72% to 62%. In 1977 House and Neuberg[5] proposed a manually phonetic transcribed data based method. In his method they first derived phonetic labels from phonetic transcription and then HMM was trained on these labels.

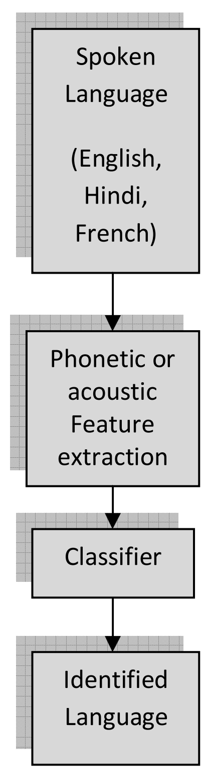

Figure 1 Block Diagram of LID System

Their method was successful in discriminating and identifying eight different languages with the help of phonotactic information. In 1980 Li and Edwars published a paper[6] in which they proposed that if the HMM model proposed by house and Neuberg is applied to real speech data then better result can also be obtained. In their method, Two statistical model were computed .The first model was based on segments and the other was based on syllables. This method was successful in achieving 80% correct identification with five languages.

Cimarusti and Ives presented a method [7] in 1982 which was based acoustic features of the speech. Autocorrelation coefficient, log area ratio, filter coefficient, cepstral coefficient and formants frequencies are some of the LPC derived feature vector used in this method. Phonetic segments was not used in this method. This method achieved 84% accuracy in language identification on eight different languages. Foil in 1986 [8] used two approaches for language identification system. First approach was based on extracting seven prosodic features from pitch and energy contour of speech while the second approach was based on extracting the formants frequencies. These features represents the characteristics of sound pattern in different languages. For classification, k-mean clustering and vector quantization method were used in this system.64% accuracy was achieved in this method.

In 1989, Goodman[9] modified the work of foil by adding more parameter in feature vector of speech signal along with the improvement in the classifier.
In 1991 Sugiyama[10] presented a method based on vector quantization classification. In this method first the features of speech ware first derived using LPC and then LPC vector quantization classification is applied . he also pointed out the difference between taking one VQ code book per language Vs Common VQ code book. The recognition accuracy achieved by this method was 80% for unknown speech. Nakagawa in 1992[11] applied four different method i.e. Vector quantization(VQ), Continuous density HMM , Discrete HMM and GMM(Gaussian Mixture distribution model). Experiments were carried out and results were obtained for four different languages. The accuracy for HMM nd GMM were 81% and VQ and Discrete HMM were 47.6%.

Muthusamy[12] in his dissertation suggested that broad phonetics , prosodic information along with acoustic information is required for automatic language identification. He first carried out the experiments on high quality speech of four languages.The results obtained by this experiments were very encouraging and prompt him to investigate more on this approach. He also tested this approach on corpus of telephonic fluently spoken speech comprises of ten different languages[13-14]. Features were taken on pairs and triples and experiments were carried out on English and Japanese languages. Later on frequency ocurrances, segment ratio and duration were also included in this experiments. A system based on all above feature got accuracy of 48.5% on short utterances and 65.6% on long utterances. Finally he drew a conclusion that the accuracy can be improved by taking phonetic level information instead of broad phonetic.

In 1995 Yan[15] in his dissertation studied the role of acoustic, phonotactic and prosodic information for language identification system. He also introduced two model in his dissertation i.e. backward LM and context dependent duration model.He achieved 91% accuracy in 45 second segments and 77% accuracy in 10 second segment. In 1996 schultz[16] presented a language identification system which was based on vocabulary speech recognition system. In his model he adopted phone lavel and word level based identification with or without language model. he implemented both bigram LM and trigram LM. The result obtained by trigram implementation is better than the bigram implementation. He designed the word based as well as the phone based system for language identification. The result obtained for word based system with trigram modelling were 84% and for phone based system it was 82.6%. From these result it is clear that the word based system outperform the phone based system. He also suggested in his dissertation that better result were obtained if more knowledge is incorporated.

Berkling in 1999[17] presented various approaches for computing the confidence measure for LID system. He proposed three different types of confidences-first type of confidence scores all poles according to the winner. In this type, the target set consist of correctly identified utterances score. In the second type of confidence, input decide the pooling of scores i.e. the target set consist of all the score for which the input and language model correspond to the similar language. In third approach, all the winning score are arranged in a single set.In his method he applied phone recognizer and then PRLM language model to evaluate the confidence measure. Method 1 gave the best result.he also added some more feature to improve the performance of the system. In 1999 Hombert and Maddieson[18] suggested 'rare' segments based Language Identification system. Detailed description of broad phonetic classes which are rare and easy to identify are used for LID system. Navaratil in 2001[19] proposed an approach based on phonotactic and acoustic features of speech. He designed a language identification system for known language identification and for unknown language rejection. In 2002 a parallel sub-word recognition(PSWR) system for language identification was proposed by Jayram[20].Sub-word recognizer apply three steps i.e. Automatic segmentation, clustering of segments and then HMM modelling.He also suggested that this system can be used in place of conventional PPR(Parallel phone recognition system).

In 2003, Adani[21] proposed anew method which is based on computing the temporal trejectories of fundamental frequency and short-term energy for segmenting and labelling the speech signal.The speech signal is fragmented into a set of discrete units which are later on used for language and person identification. He also extracted some new features in his method.His method achieved 35% equal error rate with 30 second utterances for 12 languages.
In 2003,MIT group[22] carried out the evaluation of three approaches i.e. phone recognition, support vector machine and Gaussian mixture modelling along with the fusion of all.he also pointed out the differences and progress of evaluation of NIST in 1996 to 2003.

3. Conclusion

In last few years a lots of different algorithm and approaches have been presented in the field of automatic language identification system. There has been one problem in this field in past that the comparison of different method was not possible because of lack of standard database. Now a days Standard database such as TIMIT are available for comparing the different algorithms presented in this field. National Institute of Standard and Technology (NIST) is also available for evaluation of Language identification methods proposed all over the world.

References

[1] Ma, B., C. Guan, H. Li, and C.-H. Lee, "Multilingual Speech Recognition with Language Identification," *International Conference on Spoken Language Processing*, 2002, pp. 505-508.

[2] Dai, P., U. Iurgel, and G. Rigoll, "A novel feature combination approach for spoken document classification with support vector machines," *Multimedia Information Retrieval Workshop,* 2003, pp.1-5.

[3] Leonard, R.G., Doddington, G.R. Automatic Language Identification. *Technical Report RADC-TR-74-200*, Air Force Rome Air Development Center, August 1974.

[4] Leonard, R.G. Language Recognition Test and Evaluation.. *Technical Report RADCTR-80-83*, Air Force Rome Air Development Center, March 1980.

[5] House, A.S., Neuberg, E.P. Toward Automatic Identification of the Languages of an Utterance: Priliminary Methodological Considerations. *Journal of the Acoustical Society of America*, 62(3), pp. 708-713, 1977.

[6] Li, K.P., Edwards, T.J. Statistical Models for Automatic Language Identification. *Proc.ICASSP'80*, pp 884-887, April 1980.

[7] Cimarusti, D., Ives, R.B. Development of an Automatic Identification System of Spoken Languages: Phase 1. *Proc. ICASSP'82*, pp. 1661-1664, May 1982.

[8] Foil, J.T. Language Identification Using Noisy Speech, *Proc. ICASSP'86*, pp. 861-864, April 1986.

[9] Goodman, F.J., Martin, A.F., Wohlford, R.E. Improved Automatic Language Identification in Noisy Speech. *Proc. ICASSP'89*, pp. 528-531, May 1989.

[10] Sugiyama, M. Automatic Language recognition using acoustic features. *Proc. ICASSP'91*, pp. 813-816, May 1991.

[11] Nakagawa, S., Ueda, Y., Seino, T. Speaker-independent, Text-independent Language Identification by HMM. *Proc. ICSLP'92*, pp. 1011-1014, October 1992.

[12] Muthusamy, Y.K. *A Segmental Approach to Automatic Language Identification*. PhD thesis, Oregon Graduate Institue of Science and Technology, October 1993.

[13] OGI Multi Language Telephone Speech. *www.cslu.ogi.edu/corpora/mlts/*, January 2004.

[14] Muthusamy, Y.K. The OGI Multi-language Telephone Speech Corpus. *Proc. ICASSP'92*, pp. 892-897, October 1992.

[15] Yan, Y. *Development of an Approach to Language Identification Based on Languagedependent Phone Recognition*. PhD thesis, Oregon Graduate Institue of Science and Technology, October 1995.

[16] Schultz, T., Rogina, I., Waibel, A. LVCSR-Based Language Identification. *Proc. ICASSP'96*, pp. 781-784, May 1996.

[17] Berkling, K., Reynolds, D., Zissman, M.A. Evaluation of Confidence Measures for Language Identification. *Proc. Eurospeech'99*, vol. 1, pp. 363-366, September 1999.

[18] Hombert, J.M., Maddieson, I. The Use of 'Rare' Segments for Language Identification. *Proc. Eurospeech'99*, vol. 1, pp. 379-382, September 1999.

[19] Navrátil, J. Spoken Language Recognition–A step Toward Multilinguality in Speech Processing, *IEEE Trans. Speech Audio Processing*, vol. 9, pp. 678-685, September 2001.

[20] Jayram, A.K.V.Sai, Ramasubramanian, V., Sreenivas, T.V. Automatic Language Recognition Using Acoustic Sub-word Units. *Proc. ICSLP'02*, pp. 81-84, 2002.

[21] Adami, A.G., Hermansky, H. Segmentation of Speech for Speaker and Language Recognition. *Proc. Eurospeech'03*, pp. 841-844, September 2003.

[22] Singer, E., Torres-Carrasquillo, P.A., Gleason, T.P., Campbell, W.M., Reynolds, D.A. Acoustic,Phonetic,and Discriminative Approaches to Automatic Language Identification. *Proc. Eurospeech'03*, pp. 1345-1348, September 2003.