A Study on the Selection of Speech Feature and Model Structures for Speech Recognition

李坤旺、李立民

E-mail: 9607417@mail.dyu.edu.tw

ABSTRACT

In this paper we try to figure out the best selection of speech feature and structures of hidden Markov models (HMM) to improve speech recognition accuracy. In the investigation of HMM structures we compared the performance of the whole-word, initial-final, and context dependent phoneme models. In the whole-word model structures, we study the effects of short pulse and silence models in a sentence. We adjusted the model state number to improve the recognition performance. We used the word internal tri-phone model to better represent the context affected sound in initial-final model structure. In the phoneme model, we use cross word context dependent tri-phone to precisely characterize speech sound. The decision tree algorithm was used to cluster tri-phone model states so that many of them can share the same parameters and training data. The decision tree clustering threshold was adjusted to get the best speech recognition performance. At last, we adjust parameters used in the extraction of mel-frequency cepstral coefficients. The frame length and the number of triangular band-pass filters were changed to get better performance. Experimental results show that the speech recognition performance can be significantly improved by employing these approaches.

Keywords : speech recognition ; threshold value ; triangular band-pass filter

Table of Contents

REFERENCES

[1] Bogert, B. and Healy, M. and Tukey, J. “ The Quefrency Analysis , ” Proc. Symp. On Time series Anslysis, New York, Wiley, J. pp.209-243, 1963.

[2] Davis, S. and Mermelstein, P. “ Comparing of Parametric Representations for Monosyllable Word Recognition in Continuously Spoken Sentence, ” IEEE Trans. On Acoustic, Speech and Signal Processing, pp.357-366, 1980.

[3] Beulen, K. and Ney, H. “ Autumatic Question Generation for Decision Tree Based State Tying, ” In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp805-808 Vol.2, 1998.
[4] Baum, L. E. and Petrie, T. and Soules, G. R. and Weiss, N. “ A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, ” Ann. Math., Stat., vol. 41, no.1, pp 164-171, 1970.
[5] Levinson, S. E. and Rabiner, L. R. and Sondhi, M. M. “ An Introduction to the Application of the Theory of Probabilistic Function of a Markov Process to Automatic Speech Recognition, ” The Bell System Technical Journal, Vol.62, No.4, April 1983.
[6] Viterbi. A. J. “ Error bounds for convolutional codes and an asymptotically optimal decoding algorithm, ” IEEE Trans. Information Theory, IT-13:260-269, 1967.
[7] Huang, Alex Acero, Hsiao-Wuen Hon: “ Spoken Language Processing: A Guide to Theory, Algorithm and System Development Spoken language processing, ” Prentice Hall, 2001.
[8] Enrique Vidal and Andres Marzal, “ A Review And New Approaches For Automatic Segmentation Of Speech Signals, ” in "Signal Processing V: Theories and Applications". Eds. L.Torres, E.Masgrau, M.A.Lagunas. Elsevier, 1990.
[9] Itakura, F. and Saito, S. “ Analysis Synthesis Telephony Based on the Maximum Likelihood Method, ” Proc. 6th Int. Congress on Acoustics ,Tokyo , Japan, 1968。
[10] Svendsen, T. FK Soong, “ On the Automatic Segmentation of Speech Signals, ” ICASSP, pp. 77-80, 1987.
[11] Thomas F. Quatieri, “ Discrete-Time Speech signal processing principles and practice, ” Prentice-Hall PTR, 2000 [12] Sagayama, S. and Itakura, F.: “ On Individuality in a Dynamic Measure of Speech, ” Proc. ASJ Spring Spring Conf. 1979, 3-2-7, pp. 589-590, June 1979 [13] Steve Young, “ The HTK book, ” Version 3.2, Cambridge University Engineering Department, 2002.
[14] Rabiner L. R. and Schafer, R. W. “ Digital Processing of Speech Signals, ” Prentice- Hall Inc., 1978.
[15] Boll, S. F. “ Suppression of acousitic noise in speech using spectral subtraction, ” IEEE Trans. ASSP, Vol.ASSP-27,NO.2,pp.113-120, Apr. 1979.
[16] ATAL, B.S. (1985) “ Linear Predictive Coding of Speech, ” in F. FALLSIDE - W.A. WOODS (Eds) Computer Speech Processing. Englewood Cliffs, N.J. : Prentice Hall Intenational pp. 81-124.
[17] Markel, J.D. and Gray Jr, A.H. “ Linear Prediction of Speech, ” Springer-Verlag, New York, 1976.
[18]鄭智寬, “ 語音特徵抽取方法對連續音辨認影響之研究, ” 碩士論文, 私立大葉大學電信工程研究所, 彰化, 2004,6月 [19]李佳蒨, “ 以高階隱藏式馬可夫模型作語音辨認之研究, ” 碩士論文, 私立大葉大學電信工程研究所, 彰化, 2005,6月 [20] Lee L.M. and Lee J.C. “ A Study on High-Order Hidden Markov Models and Applications to Speech Recognition, ” IEA/AIE 2006, Springer Lecture Notes in Artificial Intelligence, vol.4031, pp.682-690, Jun. 2006.