Research on speech recognition algorithm based on HTK toolbox

Lei Wang[1, a]

[1]School of Jianqiao University, Shanghai201316, China;

[a]gench_wl009@126.com

Keywords: EmbeddedSpeech recognitionalgorithm

Abstract. Speech recognition technology is a allow the machine to understand human language and speech recognition algorithm is to achieve this function, the system by using HTK toolbox to design a speech recognition system, and in-depth research and Simulation of correlation algorithm, has completed the speech recognition test, the overall recognition rate reached 98%, to achieve the desired results, reflecting the higher practical application value.

Introduction

Speech recognition technology is a kind of allow the machine to understand human language, machine through the recognition of human language and translate it into machine language machine can understand. According to the machine language to make the appropriate response and feedback, thus greatly improve the efficiency of human-computer interaction, and for the smart home, smart industrial network application scene provides a better platform for interaction.

Embedded speech recognition system is an important branch of speech recognition system, the system can provide small vocabulary speaker independent speech recognition function, and has the advantages of small size, low energy consumption, high reliability is widely used in intelligent toys, smart home, intelligent home appliances, and other remote area of Internet applications, embedded devices are generally for a particular practical application and the design and manufacture, and thus more pursuit of speech recognition accuracy and stability[1].

Introduction of speech recognition algorithm

At present, the most commonly used speech recognition algorithm a total of four, namely, dynamic time warping (DTW) algorithm, hidden Markov model (HMM,), VQ (Vector Quantization) and artificial neural network algorithm of artificial neural networks (ANN).

Dynamic time warping (DTW) algorithm is the key idea is the speech signal even to twist, stretch and deformation. It is consistent with the length of the reference model and achieve alignment feature, and continue to do two distance vector minimum matching path calculation, calculation to obtain the minimum warping function. Because each person in the voice output even if it is the same word, repeated several times when the pronunciation will be different, such as duration is different, the pronunciation of different tones, which will affect the accuracy of identification. Therefore early purely using linear time warping will be unable to meet the requirements of speech recognition, in order to change this situation, in the 1960s have Japanese scholars Itakura proposed the concept of dynamic time warping algorithm, DTW is a nonlinear warping technology, mainly the time warping and distance calculation together.

Hidden Markov model (HMM) is a kind of through the use of a large amount of speech data for training, based on a statistical model, because the statistics of a large amount of speech data, so the accuracy and recognition rate are very high. The model is evolved from the Markov chain, which is mainly used to describe the probability model of speech characteristics in stochastic process. The common topological structure of the model include: the whole connection model, the model of non spanning from left to right, and the model of spanning from left to right and parallel from left to right.

Vector quantization (VQ) is a data compression and coding techniques applicable to small vocabulary table and isolated word speech recognition, vector quantization is will a number of

scalar data combined constitute a vector. Then the vector division, given the overall quantization, to achieve data compression without loss of information. The technique consists of an encoder, a decoder and a code book. In order to get the optimal quantizer, the key is the codebook design and determine the codeword search algorithm, to solve the general approach to this problem is to use by the Lide and buzuo, gray three proposed LBG algorithm [2-3].

Artificial neural network (ANN) algorithm is refers to by the principle of mimic human neural network activity, with fast speed, strong self learning ability and adaptive characteristics, especially for in the speech recognition system, but due to the need for large amounts of training time and requires a large amount of speech data for training, so it is not commonly used, only in the laboratory were tested.

Application of speech recognition based on HTK toolbox

HTK (Toolkit HMM) is developed by the University of Cambridge, a hidden Markov model toolbox, its main role is to HMM as a model to identify the speech signal analysis and processing. The experimental kit can be adapted to a variety of speech recognition models, such as small vocabulary, isolated word, non - specific person, large vocabulary, continuous speech, specific person, etc.. HTK toolkit all the source code is open, users can to be modified according to different algorithms, to adapt to the needs of the model and algorithm, and the tool is a cross platform system, which can used on windows, can also be used on Linux or UNIX. The tool mainly includes library files and supporting tools, all tools need to use the command line mode input, the general user in use will be used with the Perl scripting language.

HTK can be divided into pre - processing tools, training tools, identification tools and results evaluation tools, as shown in figure 1.

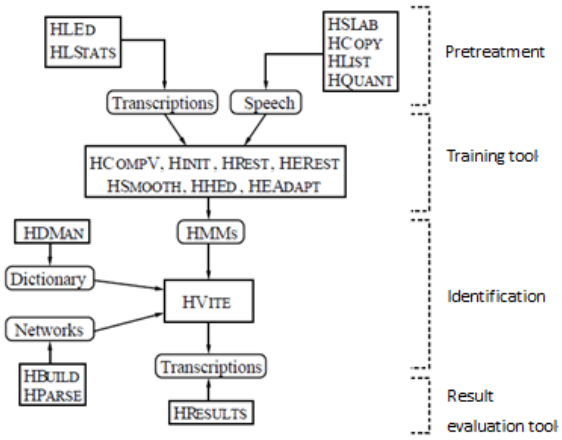

Fig. 1 HTK tools

In the use of data preprocessing and preparation tools, tool, the main function is for voice recording, mark, feature parameter extraction steps and main function includes (1) the use HSLab for voice recording, label operation. HSLab is a voice tag editor, you can add the label on the voice waveform, delete and other basic operations. (2) HLEd is a functional tool for data annotation. The tool is a simple tag file editing tool, which can be combined into a composite sequence tag by the tool. (3) Hcopy command is the extraction of speech feature parameters, the extraction of speech feature parameters for the follow-up work to be prepared. Training tools, mainly including the function a (1) HcompV tools mainly calculation a set of training data, the global mean and covariance, used to initialize the HMM model; (2) HERest is used to estimate the parameters of a single set of HMM, or is a linear transformation, it is mainly to the initial HMM parameter estimation, and can be according to the size of the amount of data for block processing, parallel processing is realized, so as to improve the training process. The most important tool in the identification tool is Hvite, through the probability statistics theory, the training of the reference model will be trained, the output of the corresponding model of the statement. Results evaluation tool, mainly refers to the Hresults tool, the tool is the HTK performance analysis tool, is the output

of the speech recognition analysis of the results, in the results will show the percentage of speech recognition accuracy.

This topic uses the HTK tool to carry out the identification of the HMM model, the specific operation process is as follows:

(1) first of all the preparatory work, including the recording of the voice files, HTK working environment, etc.;

(2) using Hcopy to extract the feature parameters of speech, that is, the key features of the speech signal code, the operator interface as shown in figure2.

(3) using the HERest tool for speech recognition training, and through the training of the establishment of a voice reference model library. The training results are shown in figure 3.

(4) speech recognition using tool Hvite, through the comparison of the new reference model library, for speech recognition than the highest similarity of the results of the output.

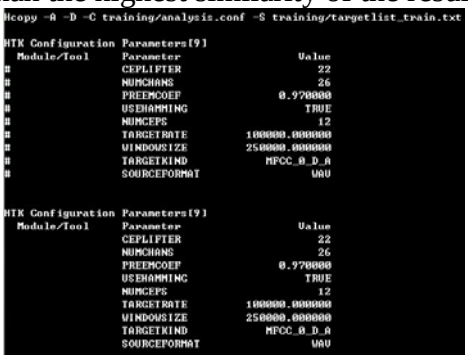

Fig. 2HTK extraction of characteristic parameters

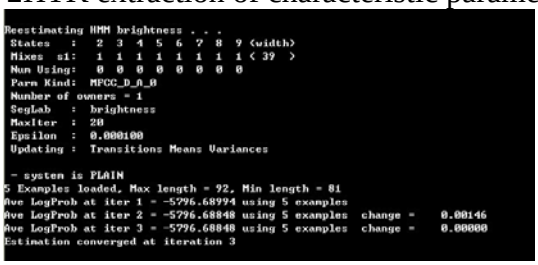

Fig. 3Speech recognition training

(5) the output operation of the speech recognition result using the HERest command, because the number of speech samples is not much, so the recognition rate is 98.67%, as shown in figure4.

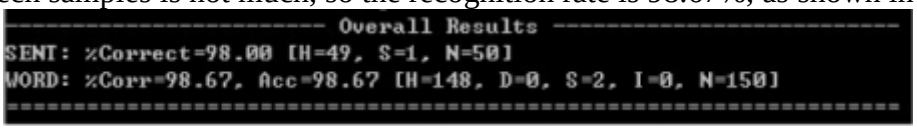

Fig. 4Speech recognition results

Concluding remarks

Speech recognition technology for human-computer interaction to create a new mode of communication, and speech recognition correct rate and the recognition efficiency is determined by the algorithm, the system based on HTK realized the HMM algorithm research, create the good preconditions for the realization of the embedded speech recognition system.

Reference

[1] Jing Jiamin, Liu Jia, Liu Runsheng, the application of speech recognition technology based on [J] in embedded system HMM, computer application, 2003, (10): 12-13

[2] Alba E,Dorronsoro B. The exploration/exploition tradeoff in dynamic cellular genetic algorithms[J].IEEE Transactions on Evolutionary Computation，2005，9(2):126-142

[3] Xie Jinhui.HMM and its application in speech processing [M]. Wuhan: Central China University of science and Technology Press, 1995,110-113