The application of Speech Recognition Technology based on HMM

Guilin Yan

Department of Foreign Languages, Yun yang Teachers College, Hubei, China

37545473@qq.com

Keywords: Hidden Markov Model; Speech recognition; Viterbi Algorithm.

Abstract. It was a great improvement for the speech signal digital processing technology to use the model of Hidden Markov Model in 1980's, which is used to describe the speech signal. In this paper, with the help of the speech recognition technology and the interpretation of Hidden Markov model, by means of the analysis on HMM and Viterbi algorithm, it discusses the speech recognition based on HMM.

Introduction

At present, with the continuous development of modern science and technology, the society of human beings has entered the information age. Information technology has began to infiltrate into every aspect of people's life, at the same time, the practical application of speech recognition technology can help people to experience the operation if machines, which can be very convenient for people to control machines through using natural language to implement the intention of people and completely replace the pattern that people used hands or other body parts to operate machines. The appearance of voice controlling mode of operating machines makes people more convenient, and safer, as well as with more humanized operation experience. Therefore, the research of speech processing technology can make the speech signal more efficiently generate, store, transmit and recognize and so on, which has become a hot issue of current research on the processing of speech signal.

The Interpretation of Speech Recognition Technology

Speech recognition technology is mainly to meet the demand of normal communication between human and machine, which is a kind of high-tech information technology through human voice commands to control the machines and help people to perform the expected work. During the process of the technical application of speech recognition, speech recognition is generally divided into two phases: the first phase is the system of "learning" or "training" process. The main task of this phase is to adopt some models of speech training and learning with the known speech template and set up the completed training language model of speech signal data that is based on the training and learning. The second phase is the "recognition" or "test" phase of the speech signal. After pre-processing of the unknown voice data and having feature extraction, inputting the information into the speech model that has been trained, then, according to a certain decision comparison criteria, compared with the speech template. Making data analysis and identification for the unknown speech signal on the established speech model, then it can withdraw the final recognition results of the unknown speech signal. The basic process of voice recognition is shown in Fig. 1:

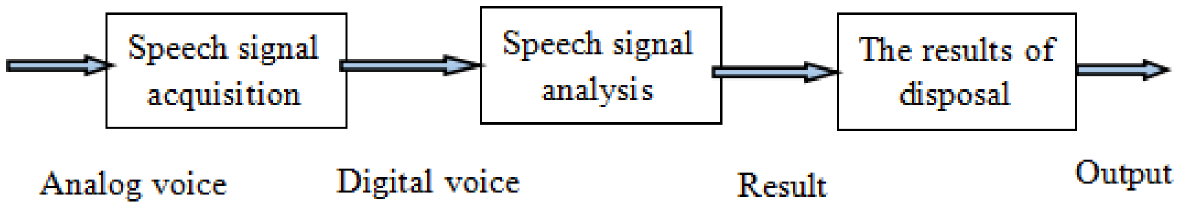

Fig.1 The Flow Chart of Speech Recognition

Interpretation of HMM

Hidden Markov Models (HMM) is used to simulate the changes of the statistical characteristics of the signal by using Markov chains, while this kind of change can be described indirectly through the observation of the sequence, the observed events and the states are associated with a set of probability distribution, which is a probability model of double random process. The first layer is Markov chain, it can describe the transferring of the state, which is the basic stochastic process. Another layer can describe the corresponding relationship between the descriptive statistics of each state and the observation value. Because the observed event and the state does not correspond one to one, it can not directly see the state, but it can experience the existence and properties of the process to perceive the state randomly, thus, it is so called "Hidden Markov Model, namely, HMM.

HMM and Viterbi Algorithm

The speech recognition system based on HMM can include two sub-processes: the training process and the scoring (recognition) process. In the training phase, the system adopts a series of training parameter to observe estimation of the HMM, setting up HMM model for each word in the glossary. In the recognition phase, it can calculate out the probability score of data as well as each word model, then select the maximum probability score of word model as the recognition results. In this process, the Vietbri algorithm is used to compare the HMM model of human speech transmission sequence with the memory, which can effectively find out the best matching result. Because of the limitation of the computational complexity, as for the real-time speech recognition that is based on HMM, it needs to design an efficient hardware architecture to implement the vietbri decoding process, which can accelerate the process of recognizing HMM. In the definition of HMM, it includes a state transition matrix: *A*; an observation value output probability: *B*; as well as a probability of an initial state. As for the given multi-dimensional observation value sequence (which is also known as feature vector). And the expression of HMM in $\lambda = (A, B, \pi)$, which can calculate the probability of the observation value sequence $P\ (o/\lambda)$. In HMM, word acts as a unit, the recognizer can calculate and compared with all $P\ (o/\lambda_V)$ $(v=1, 2... W)$, among them, W is the number of word model.

Adopting the the minimum route of log probability and state probability respectively, transform the traditional vietbri algorithm, thus the calculation process of $P\ (o/\lambda_V)$ can be shown as follows:

Initialization: When t=1, $1 \leq j \leq N$

$$\delta 1\ (j) = \pi j + \log bj\ (o1) \quad (1)$$

Recurrence: When $2 \leq t \leq T$, $1 \leq j \leq N$

$$\delta t\ (j) = \min_{i=j-1} [\delta t\text{-}1(j) + aij] + \log bj(0t) \quad (2)$$

3 End: When t=T

$$P\ (o/\lambda v) = \min_{1 \leq i \leq N} [\delta t\ (i)] \quad (3)$$

In the formula, a= {aij} is the state probability from the state i to state j, N is the numb of state, T is the number of the frames of the feature vector of $O= [O_1, O_2... O_T]$.

Adopting directly the adder, comparator of FPGA as well as the logic operation to achieve the formula (2) and formula (3), which can improve the efficiency of the system greatly. The structure of the system can be shown in figure 2.

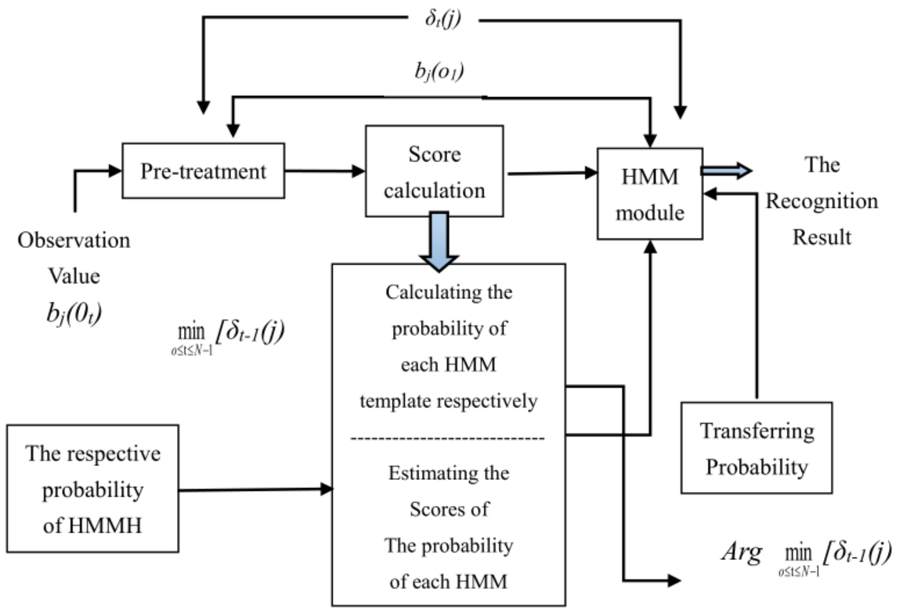

Fig. 2 Decoding of Viterbi

The Recognition of HMM

The basic task of HMM recognition is to recognize the unknown input speech signals. First of all, through the endpoint detection, it can acquire the correct speech segments, and then through the same feature extraction with the training process, it can get the observation sequence o=o1, o2,... oT, using forward -backward algorithm or Viterbi algorithm, so as to calculate the observation value sequence o=o1, o2,... oT, as well as P (o/ λ i) under the condition of λi(l≤i≤r) in each trained HMM models, taking the maximum probability of the model as identifying entry.

Design function void diagnosis (HMM HMM [F], double O), under the condition of the known trained HMM model hmm [l], hmm [2]… hmm [F], after extracting the unknown voice signal through feature extraction of MFCC, recognizing HMM with the observation of the acquired value sequence, o= {o1, o2... OT}, finally showing the results of speech recognition directly to users.

Initialize the maximum output probability: max=o, at the same time, the parameter index of the corresponding model of the maximum output probability: d=o;

Adopt Viterbi algorithm, calculate the output probability of each model by loop computation, make maximum output probability as well as the parameter index of the corresponding model of the maximum output probability to be assigned with max and d respectively;

```
For (i=l; i≤N; i++)
{
//Calculating the output probability by using Viterbi algorithm;
//Calculating the output probability by using Viterbi algorithm;
P=Viterbi (hmm[i], o);
If (p>max)
{
//Finding out the maximum probability;
Max=p;
d=i;
}
}
```

The matching result is model with hmm[d], namely, it can recognize the speech of the D th model correspondingly.

Conclusion

By means of the research and experiment of speech signal recognition method based on HMM, it shows that putting HMM model into the application of speech recognition successfully has bright prospects. It is believed that after further study and exploration on HMM model, the method of feature extraction, the endpoint detection of speech as well as many other aspects, the function and the performance of speech recognition system can be further improved.

Reference

[1] Yoshizawa, Shingo, Miyanaga, Yoshikazu. 2004, A High-Speed HMM VLSI Module with Block Parallel Processing. Electronics and Communications in Japan, Part III: Fundamental Electronic Science, vol.87, p.26-28.

[2] Salazar Jaime, Robinson Marc. 2003,A Hybrid HMM-Neural Network with Gradient Descent Parameter Training. Proceedings of the International Joint Conference on Neural Networks, vol.32, p.1086-1091.

[3] Tze Fen Li. 2003, Speech recognition of mandarin monosyllables. Pattern Recognition, vol.36, p. 2713-2712.

[4] R.Hoffmann, C.-M.Westendorf. 1997, The development of analysis methods for speech recognition. Behavioural Processes, vol.39, p.113-125.

[5] H.Hennanski.et al, 1985, Perceptually based linear predictive analysis of speech, ProcIEEE.Acvust. Speech Signal Process.Vol.1, p.33-36.