G-20

HTK を用いた背景雑音環境下における音声認識

Speech recognition under the conditions of background noise using hidden Markov model toolkit

○清水秀樹[1]
*Hideki Shimizu[1]

Abstract: The hidden Markov model toolkit (HTK) is a toolkit that can perform speech recognition using hidden Markov models. In this paper, we report the simulation results of speech recognition under the conditions of noise using the HTK.

１．概要

現在，音声認識技術は携帯通信機器やカーナビゲーションといった，様々な分野で利用され，日常生活をよりよく過ごす際に必要とされる手段である．また，雑音環境下における様々な認識手法が提案されつつある[1].

本論文では，隠れマルコフモデル（Hidden Markov Model : HMM）を利用した HTK を用い，背景雑音下での数字音声認識実験結果について報告する．

２．音声認識

音声認識とは，人間の発する音声をコンピューター上で認識させ，それを文字に変換する技術である．音声認識を行うためには主に，MFCC（Mel-Frequency Cepstrum Coefficients）や HMM が用いられている．

３．MFCC

一般に，音声認識の音声の特徴量として MFCC が広く用いられている．MFCC を得るためには，まず，与えられた音声信号をディジタル化し，FFT（高速フーリエ変換）を行い，振幅スペクトルを求める．次に，振幅スペクトルに対してメル帯域化，および，その対数をとった後に DCT（離散コサイン変換）を行い，ケプストラム係数を求める．そして，ケプストラムに高次の情報を削除するリフタリングを行い，求められた低次の情報が MFCC となる[1].

４．HMM

HMM は，時間的に変化する性質を持つものに対して有効的な確率モデルとして知られ、現在では代表的な音声認識手法の一つとして広く用いられている[2].
HMM では，各内部状態に割り当てられた，次の状態に遷移する確率，自分に戻ってくる状態に遷移する確率，および状態内の特徴ベクトルの確率を学習が行われる．また，一般的に音声認識では左から右へと遷移するモデルが用いられる．

さらに，出力系列が与えられたときに，どのような状態遷移が行われてきたかが隠れているため，隠れマルコフモデルと呼ばれる．

５．HTK

HTK(Hidden Markov Model Toolkit)は，HMM の構築・学習・認識・評価などといった操作を行うツール群である[2]. 本研究で使用した HTK のコマンドは Table 1 の通りである．

コマンド名	機能
HCopy	特徴抽出
HInit	HMM の初期化
HRest	HMM の学習
HParse	文法記述をネットワーク表現に変換
HVite	ビタビアルゴリズムによる認識
HResults	認識結果の集計

Table 1．HTK の基本コマンド

６．HTK の使用手順

まず，HCopy コマンドで音声データの MFCC を抽出し，MFCC のデータを得る．また，MFCC のデータを用いて，HInit コマンドによる HMM の初期構成の設定を行う．

次に，HRest コマンドにより，HInit コマンドにて初期化された HMM に対し，Baum-Welch アルゴリズムによる学習をパラメータの変化量が閾値以下になるまで繰り返し行う．このような HMM の学習後，認識を行う．

なお認識を行う前に，あらかじめ単語系列規則を示した文法ファイルを作り，HParse コマンドにより，認

１：日大理工・学部・数学

識を行うための HMM ネットワークを作成した上で，認識と評価を行う．

認識はネットワーク内の各状態がどの HMM に対応するか定義した単語辞書ファイルの作成した後，HVite コマンドでビタビアルゴリズムを実行し行う．また，認識結果と比較するための正解リストを作成した後，HResults コマンドにより，認識率の評価を行う[2].

７．予備実験

本研究における予備実験では，無音環境下での単数字 0~9(/zero/, /ichi/, /ni/, /san/, /yon/, /go/, /roku/, /nana/, /hachi/, /kyuu/)を 1 人の発音者につき，それぞれ 3 回ずつ録音し，HTK を用いて，HMM の学習，認識を行った．

実験では，録音した音声データを wav データとして用意し，WaveSurfer[3]ソフトを用いて，音声データ内の何秒から何秒までにどのようなデータが入っているかの情報の付与をするラベル付けを行った．ラベル付けは，「無音部分(sil)＋数字(zero)＋無音部分(sil)」のように行い，0~9 のそれぞれ 3 回ずつ計 30 個の録音データ全てに対しを行った．

ラベル付け後, 全ての音声データの特徴抽出, sil, 0~9 の HMM の初期化と学習，ネットワークの作成，認識および評価を行ったところ，認識率は 73.33%となったが, 評価を行う際に「無音部分(sil)」を含まない「数字」のみの評価を行うようにしたところ，認識率は 100.00%に向上した．

このような結果が得られたのは，正解の集計を行う際に影響が出たため, 無音部分(sil)は必要ではなかったものと考えられる．

次に述べる背景雑音下における音声認識実験では，雑音下における HMM の性能についての実験であるため，予備実験の結果と考察から，音声データの「有声部分」のみを用いた場合の認識性能の評価を行うこととした．

８．背景雑音下における音声認識実験

本実験では予備実験と同様に単数字 0~9(/zero/，/ichi/, /ni/，/san/，/yon/，/go/，/roku/，/nana/，/hachi/，/kyuu/)を用いて，背景雑音が付加された場合の音声認識を HTK を用いて行う．背景雑音の種類として，音の大きさ，高さが定常的な雑音（運転中の車の車内など）と非定常的な雑音（電車が通るときの線路の高架下など）が付加された場合での録音を 0~9 のそれぞれ 3 回ずつ行う．

なお，予備実験同様，音声データを WaveSurfer ソフトでラベル付けを「無声部分(less)＋数字(zero)＋無声部分(less)」のように 0~9 の音声データ全てについて行う．ラベル付け後，HTK を用いて特徴抽出，HMM の初期化と学習，ネットワーク作成，認識および無声部分を含まない場合の評価を行う．

次に，雑音下における音声認識実験に際して，背景雑音を含んだ音声のみでの評価，予備実験で用いた無音環境下での学習済み HMM を用いて背景雑音入りの音声データの認識率の評価を行い，背景雑音がどのような影響を及ぼすのか考察する．

以上述べた雑音下音声認識実験の結果について，学術講演会当日に報告する予定である．

９．参考文献

[1] 荒木雅弘：「イラストで学ぶ音声認識」，講談社，pp.2-12,60-71，2016 年 7 月．
[2]荒木雅弘：「フリーソフトでつくる音声認識システム」，森北出版株式会社，pp131-165，2017 年 4 月．
[3]https://sourceforge.net/projects/wavesurfer/ : (WaveSurfer のダウンロード先).