P-8

CNC を用いた雑音下における音声認識

Speech Recognition Under Noisy Conditions Using A Cascaded Neuro-Computational Model

○会沢　純将[1], 保谷　哲也[2]
Yoshiyuki Aizawa[1] and Tetsuya Hoya[2]

Abstract: Recently, the application of artificial neural networks to the speech recognition area has attracted a growing interest. In this direction, cascaded neuro-computational model (CNC) was proposed as a psycholinguistic model of human speech recognition, whereas it handles real speech input unlike previously proposed models. In this paper, we report the simulation results of spoken digit recognition performed under noisy situations using CNC.

１．概要

現在,音声認識システムの研究では, 統計的手法である隠れマルコフモデル (Hidden Markov Model : HMM) が広く用いられている. 本研究では, 人間の心理学的アプローチに基いたモデルである CNC を用いて, 雑音下における数字音声認識を行う. 学術講演会当日では, その実験結果について報告する予定である.

２．音声認識

音声認識とは, 一般に, 音声信号のパターン認識を指す. また, 人間は外界から得た情報から, 認識に必要なデータを抽出し, 脳内のデータベースとパターンマッチングを行い, そのマッチングした情報を, 音や文字など意味のある情報として認識していると考えられている. このような人間の脳内で日常的に行われている情報処理をコンピュータで実現することが, 今日に至るまでに大きな目標の一つとなっている.

３．Mel-Frequency Cepstrum Coefficients

一般に, 音声認識では,音声特徴量として Mel-Frequency Cepstrum Coefficients (MFCC) が主に利用されている. 通常, MFCC はマイクなどにより入力された音声信号をフーリエ変換(FFT)して得られるスペクトル情報に対してフィルタバンクによる分析を行った後, 離散コサイン変換(DCT)を行うことで求められる.

４．Cascaded Neuro-Computational (CNC) Model

CNC は, 心理言語学的アプローチによる人間の音声認識のメカニズムに基いて構成された人工ニューラルネットワークである[1]. CNC は, Figure 1 に示されるように, 第１層には受容ユニット群, 第２層には単語候補ユニット群, および第３層には単語ユニット群を有するような３層構造を成している.

より具体的には、まず、音声の特徴抽出を MFCC 形式により行い, フレーム単位の入力として第１層ユニット群に与える. 第１層は, その与えられた MFCC により, ラディアル関数 (RBF) によって実現される受容ユニットから構成される. ここで h_i を i 番目の RBF ユニット, x を MFCC からの入力データ, c_i をセントロイドベクトル, σ を半径とすると, その出力は次式 1 のように表される[1].

$$h_i = exp\left(-\frac{\|x-c_i\|_2^2}{\sigma^2}\right) \quad (1)$$

学習時には, パターンマッチングを行い適宜, 各層のユニット追加が行われる. 第２層では各フレームを第１層に入力した際に得られる最大発火ユニットの時系列データ入力により, 単語候補の出力が得られる. 第３層では, 第２層からの出力を各単語ごとに集計し, 最大発火した単語ユニットを出力とする.

CNC の特徴として, このような第１－３層からなるネットワーク構造は事前に用意されたものではなく, 入力されたデータによって, 順次自己構造化されていき, ネットワーク全体を再学習させることなく, 新たにデータを追加させることが可能である[1].

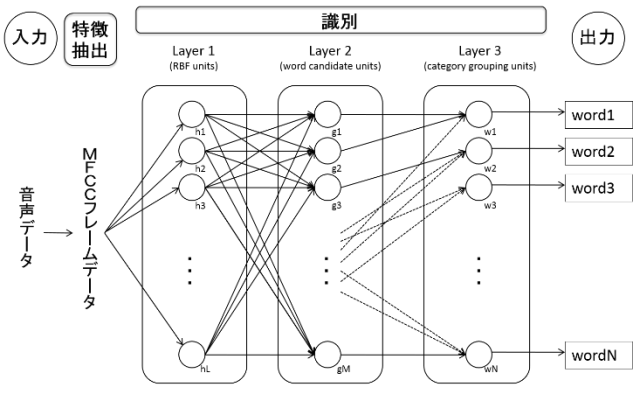

Figure 1. CNC

１：日大理工・院(前)・数学　２：日大理工・教員・数学

５．HMM

本研究では, CNC との比較および検証実験のため, 現在の音声認識の主流である HMM を使用する.

HMM は, 時間的に変化する性質を持つ問題(音声認識や画像認識)に対して有効な統計学習モデルであることが広く知られている[2]. HMM では, 各状態における次の状態遷移する確率と自己ループする確率, および状態内での特徴ベクトルの確率が学習により決定される. 一般に音声認識においては, 左から右へと遷移していくモデルが扱われる (Figure 2) .

また, ある出力系列が与えられた時に, どのような状態遷移が行われたかが隠されているため, 隠れマルコフモデルと呼ばれる[3].

本研究では, HMM に関しては, HTK(Hidden Markov Model toolkit) を用い, 特徴抽出から実験の評価まで一括して行う.

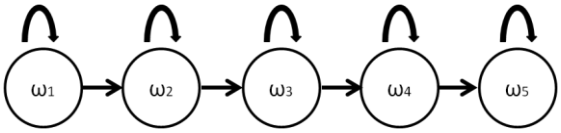

Figure 2. HMM（left-to-right モデル）

６．雑音重畳数字音声認識実験

本実験では, 1 ~ 9(/ichi/, /ni/, /san/, /yon/, /go/, /roku/, /nana/, /hachi/, /kyu/)の数字を組み合わせた２１通りの日本語数字列 (Table 1) を考慮した. また, 各数字列は Table 1 に示されるような３つの数字単語からなり, それぞれ７人の話者により発音・録音された１０パターンを実験データとして用いた. その内, ６パターンを学習に使用し, 残りの４パターンを本実験の評価データとした. 実験では, ３つの単語を１つの単語(クラス)とみなした. また, 雑音音声への変換は, Octave (https://www.gnu.org/software/octave/)の randn() 関数を用いて, 正規乱数を作成した上で, 各音声に重畳し, 実験に用いた.

具体的には, 音声対雑音比を SNR とし, 音声信号を S, 雑音信号を N, 雑音重畳音声を NS としたとき, 式2, 3 より求めることができる. また本実験に際して, SNR は-5, 5, 10, 15, 20(dB)とする.

本研究では, 用意したデータを, それぞれ CNC モデルを用いて学習, 認識を行う.

学習の際は, 雑音重畳をしていない音声データを用い, 認識の際は, 雑音重畳を施した音声に対して実験を行う.

以上述べた雑音重畳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.

$$\mathrm{SNR} = 20 * \log 10 \left(\frac{\|S\|}{\|N\|}\right) \tag{2}$$

$$\mathrm{NS} = \mathrm{S} + \frac{\mathrm{N}}{\|N\|} * \frac{\|S\|}{10^{(0.05 * SNR)}} \tag{3}$$

118	123	147	222	239	355	369
416	451	456	573	642	645	745
789	867	871	898	932	967	983

Table 1. 数字列一覧

７．参考文献

[1] Tetsuya Hoya and Cees van Leeuwen : “A cascaded neuro-computational model for spoken word recognition”, Connection Science , vol. 22, pp. 87-101, February 2010.

[2] Richard O. Duda , Peter E. Hart , and David G. Stork : “Pattern Classification Second Edition”, John Wiley & Sons , pp.125-138, April 2009.

[3] 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, pp.22-23,130-147, October 2007.

[4] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄:「音声認識システム」, オーム社, pp.22-23, October 2006.

[5] Steve Young , Gunnar Evermann, Mark Gales , Thomas Hain, Dan Kershaw, Gareth Moore, Julian Odell , Dave Ollason,Dan Povey, Valtcho Valtchev, and Phil Woodland : “The HTK Book (for HTK Version 3.3)”, Cambridge University Engineering Department, April 2005.