G-22

Hidden Markov Model Toolkit を用いた単語音声認識について

Hidden Markov Model Toolkit for Word Recognition

○中田圭介 [1]
*Keisuke Nakata[1]

Abstract : HTK is a tool kit to perform construction / learning / recognition / an evaluation of HMM. In this article, I speak that I realize word recognition using HTK.

1、はじめに

人間の音声知覚は音素や単素のレベルにより分析することによって行われている．“**Hidden Markov Model Toolkit**”（**HTK**）は、**HMM**[1]の構築・学習・認識・評価などのためのツールキットである．データを準備するためのツールや評価のためのツールは主として音声認識を対象にしたものであるが,学習や認識のツールは対象に依存しないので,動画像の認識などにも応用できる．本論文では**HTK**を用いて単語認識を実現することについて述べる．

2、Hidden Markov Model Toolkit

HTK は Table 1 に示されるように７つの基本コマンドで構築されている．**HMM** の学習は **Baum-Welch** アルゴリズムで、パラメータの変化量が閾値以下になるまで繰り返す．具体的には,音声サンプルを HSLab コマンドで実験用の音声を録音し,学習に必要な正解ラベル付けをした後,**HCopy** コマンドで音声の特徴抽出を行う．次に **HMM** の構成を決める．状態数・状態遷移数・混合数などを決め,**HMM** 構成情報ファイルを作成する．そして,**HInit** コマンドを使い学習データから **HMM** のパラメータの初期値を決め,**HRest** コマンドを使って **HMM** 構成情報ファイル上のパラメータを繰り返し更新する．最後に,認識を行うための文法を書き，**HParse** コマンドを使って文法を **HMM** のネットワーク形式に変換する．認識は **HVite** コマンド,認識結果は **HResults** コマンドを用いて評価する．

コマンド名	機能
HSLab	音声の録音、ラベル付け
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

3、実験

一人の特定話者により数字の組み合わせ（データセット１）２１パターンを１パターンずつそれぞれ 10 個用意し数字ごとにそれぞれ切り抜いておく（データセット２）．そこから各パターンよりそれぞれ５個を学習データとして,残りの５個をテストデータとして用い,各数字列ではどのような認識結果が出るか単語音声認識実験を行った．

データセット１：/yon-go-roku/や/ichi-ichi-hachi/など三つの数字を組み合わせたもの．

データセット２:
/ichi/,/ni/,/san/,/yon/,/go/,/roku/,/nana/,
/hachi/,/kyu/の９つの数字×７０個.

４、実験結果

単語音声認識の結果の一部を Table 2,Table 3 に示す. Table 2 は各単語別における単語一致数と平均認識率であり,Table 3 は数字列別の詳しい出力結果表である. ある程度認識することはできたが,/yon/が加わっているパターンだけ認識率が低いことが判明した.

５、終わりに

本研究では,HTK を用いた単語音声認識について,/yon/の加わっているパターンは認識率が低かったが,ある程度認識は出来るものとして結論した. 今後の課題としては,Table 2,3 の結果から示されているように/yon/の加わるパターンの認識率向上ということが挙げられる.
また,不特定話者の場合について認識率がどのように変化するのかを実験によって確かめ、その実験結果についても学術講演会当日に発表する予定である.

	単語一致数/35	平均認識率
ichi	31	89%
ni	35	100%
san	35	100%
yon	29	83%
go	31	89%
roku	32	91%
nana	35	100%
hachi	35	100%
kyu	33	94%
全体	296/315	94%

Table ２：各単語別における単語一致数と平均認識率.

	118			123			147			222			239			355			369		
1	1	1	8	1	2	3	1	4	7	2	2	2	2	3	9	3	5	5	3	6	9
2	1	1	8	1	2	3	1	4	7	2	2	2	2	3	4	3	5	5	3	6	9
3	1	1	8	1	2	3	1	4	7	2	2	2	2	3	9	3	5	5	3	6	9
4	1	1	8	1	2	3	8	1	7	2	2	2	2	3	1	3	5	5	3	6	9
5	1	1	8	1	2	3	1	4	7	2	2	2	2	3	9	3	5	5	3	6	9
	416			451			456			573			642			645			745		
1	4	2	1	4	5	5	4	5	6	5	7	3	6	4	2	6	4	5	7	f	5
2	1	1	6	1	5	1	4	5	6	5	7	3	6	4	2	6	4	5	7	f	5
3	4	1	6	1	5	1	4	4	5	5	7	3	6	4	2	6	4	5	7	f	4
4	1	1	6	4	5	1	4	4	5	5	7	3	6	4	2	6	4	5	7	f	4
5	4	1	6	4	5	5	3	5	6	5	7	3	6	4	2	6	4	5	7	f	5
	789			867			871			898			932			967			983		
1	7	8	9	8	6	7	8	7	1	8	9	8	9	3	2	9	6	7	9	8	3
2	7	8	9	8	6	7	8	7	1	8	9	8	9	3	2	9	6	7	9	8	3
3	7	8	9	8	6	7	8	7	1	8	9	8	9	3	2	9	6	7	9	8	3
4	7	8	9	8	6	7	8	7	1	8	9	8	9	3	2	9	6	7	9	8	3
5	7	8	9	8	6	7	8	7	1	8	9	8	9	3	2	9	6	7	9	8	3

Table 3: 数字列別の出力結果.

６、参考文献

[1] 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・2007. pp.130-147.

[2] 鈴木清四郎、保谷哲也、“A cascaded neuro-computational model を用いた連続音声認識について”、平成 22 年度（54 回）日本大学理工学部学術講演会

[3] Tetsuya Hoya, Seishiro Suzuki, and Yoshihisa Ishida “Concatenated Spoken Digits Recognition Using A cascaded Neuro-Computational Model,” subject to revision for the *Neurocomputring*.

1：日大理工・学部・数