Articulatory Feature Asynchrony Analysis and Compensation in Detection-Based ASR

Author: I-Fan Chen and Hsin-Min Wang

Professor:陳嘉平

Reporter :吳柏鋒

摘要

- 簡介
- 前端發音特徵偵測(AFDT)和
  AF asynchrony分析
- 使用Condition Random Field(CRF)的
  後端音素辨識
- 實驗

簡介

- 以偵測為基礎的ASR可分為前端知識偵測器與後端知識整合器，兩個主要元件:

1. **前端知識偵測器**:收集不同的語音相關knowledge attribute形成knowledge source

2. **後端知識整合器**:將knowledge source整合成高階語音單元(例:syllables, words and sentence)

簡介

- 根據先前研究AFDT errors 和AF asynchrony為造成辨識率的下降的兩大主因
- 設計一些variable-controlled實驗，得知AF asynchrony對辨識率的影響較大
- 並提出幾個方法來改善以偵測為基礎的ASR系統的辨識率

前端發音特徵偵測(AFDT)

- 選擇GP特徵集為偵測基礎，其中含8 個AFs(“A” , “I” , U” , “E” , “S” , “h” , “H” , and N”)以及額外3個AFs(“a” , “i” , and “u”)，共11個AFs
- 除了11 AFs外，AFDT偵測TIMIT的61個音素來作AF asynchrony分析
- 以12維MFCC+energy作為類神經網路的輸入，
  其擷取25ms的漢明窗且作10ms平移，而輸出為72維向量

使用AFDT對AF和phone作alignment

- **AF 序列 alignment:**

  -每個AF個別作force-aligned，決定每個frame中特定AF的label為active或inactive

例:一長句有12 phones，對應到短序列 /n_u u n_u/ ，

n_ 表示inactive feature，則AFDT只要產生的事後機率發生錯誤，整個aligned就會錯誤，故就無法正確的作AF asynchrony分析

使用AFDT對AF和phone作校正

- 發音特徵 AF_K 的音框i的smoothed active事後機率為:

$$P_{Smoothed}^{AF_K\_active}(i) = \alpha \cdot P_{AFDT}^{AF_K\_active}(i) + (1-\alpha) \cdot P_{realignedPhoneDerived}^{AF_K\_active}(i),$$

其中 $P_{AFDT}^{AF_K\_active}(i)$ 為前端AFDT產生的事後機率,

$P_{realignedPhoneDerived}^{AF_K\_active}(i)$ 為AF_K's值不是0(inactive)就是1(active), α =0.2平移加權因子

使用CRF做後端語音辨識

- 提出以下三種方法訓練CRF模型:

1. **Oracle data training(OT):**
   使用phone-derived, frame-based AF value sequence訓練.
2. **Detected data training(DT):**
   直接由前端AFDT產生的特徵向量訓練
3. **AFDT aligned data training (AT):**
   利用phone序列中個別的音框作alignment以及作為資料訓練的11 AFs序列來訓練

實驗

- 使用**TIMIT** 語料庫，其中training set(3296句)，development set(400句)，test set(1344句)

- 以Oracle data training和Detected data training訓練的CRF模型為**baseline**

實驗

(1)**Ideal case**: AF由human phone label轉換而來
(2)**Real case:** 由AFDT自動偵測產生AFs

Table 1. The phone recognition results of the oracle data trained (OT) CRF model and the detected data trained (DT) CRF model.

Test Data Type	System	Corr	Acc
Ideal (upper bound)	OT CRF	98.31	98.28
Detected (real case)	OT CRF	70.55	34.38
	DT CRF	57.30	56.14

Acc=Corr-insertion rate

實驗

- **The AFDT aligned data trained system(AT):**

  -指出AFDT errors 和AF asynchrony,在以OT訓練的CRF模型中ideal case與real case 效能的不同

Table 2. The phone recognition results of the AFDT aligned data trained (AT) CRF model.

Test Data Type	System	Corr	Acc
Ideal (upper bound)	AT CRF	71.49	70.31
Detected (real case)	AT CRF	64.87	62.32

實驗

- **AF asynchrony補償:**
  1. AFDT使用long-term資訊作補償
  2. CRF integrator使用long-term資訊作補償
  3. 以AT訓練的CRF系統作補償

實驗

Table 3. The mean AF-phone boundary distances (in frames) of GP11 AFs generated by the original AFDT (using MFCC as input) and the AFDT using long-term information.

AFDT sys	a	A	E	h	H	i	I	N	S	u	U
MFCC	0.94	4.27	1.18	0.94	0.59	1.07	2.44	1.74	0.56	1.88	7.16
Long Term	0.75	3.57	0.94	0.77	0.38	0.91	2.36	1.15	0.43	1.32	4.55

實驗

Table 4. The phone recognition results of the AT-trained CRF system with AF asynchrony compensation

Test Data Type	System	Corr	Acc
-	CI-HMM	69.02	63.45
-	CD-HMM	75.76	65.78
Detected (real case)	OT CRF(±3)	75.24	47.97
	Long Term AFDT + DT CRF(±3)	64.58	63.12
Ideal (upper bound)	Long Term AFDT + AT CRF	74.96	73.64
	MFCC AFDT + AT CRF(±3)	72.87	71.62
	Long Term AFDT + AT CRF(±3)	76.41	74.97
Detected (real case)	Long Term AFDT + AT CRF	69.83	66.97
	MFCC AFDT + AT CRF(±3)	66.21	63.16
	Long Term AFDT + AT CRF(±3)	**71.01**	**67.67**