

ISTCにおける2006年度 音声認識ソフトウェアの開発成果

河原達也（京都大）
李 晃伸（名工大）

これまでの経過

- 1995～1997 IPSJ/SLP傘下 WG
 - JNASコーパスの設計
- 1997～2000 IPAプロジェクト
 - 「日本語ディクテーション基本ソフトウェア」の開発
- 2000～2003 連続音声認識コンソーシアム CSRC
 - 認識ソフトウェアの改善
 - 音響・言語モデル等の充実
- 2003～2008 e-Society基盤ソフトウェア
- 2003～2006 音声対話技術コンソーシアム ISTC
 - 対話システムを指向した音声認識の改善

ISTCでの主要開発目標

- 音声認識エンジンJuliusの性能改善・機能追加
→ Julius 3.5 2005年11月リリース
- 音声認識エンジンJuliusのSAPI/SALT対応
→ 2003年度
- 音声認識ソフトウェアのカスタマイズを容易に
→ 2004/2005年度
<http://htk.ar.media.kyoto-u.ac.jp/julicus/>
- 統計的言語モデルの構築を容易に
→ 2006年度
http://www.ar.media.kyoto-u.ac.jp/members/misu/tool/web_collect.tar.gz
- 多様なパッケージ
 - 英語版、雑音対応版

Julius の本年度開発成果

Rev. 3.5.1 (2006/3/31)
Rev. 3.5.2 (2006/7/31)
Rev. 3.5.3 (2006/12/19)

今年度の成果

- 性能の改善
 - ① 処理の高速化
 - ② MAP-CMNの導入
 - ③ グラフ出力の修正と改善
 - ④ 認識用文法の最適化
 - ⑤ メモリの最適化・遅延の改善
- 互換性の拡大
 - ⑥ MFCC フルサポート
 - ⑦ HTK Config の直接読み込み
 - ⑧ HTKからJulianへの文法変換
- その他
 - ⑨ ホームページ再構築
 - ⑩ バグ修正とコード整理

①処理の高速化

- 出力確率計算で全ての除算を乗算に置換
- PCにおいて計算速度の大幅な改善を確認

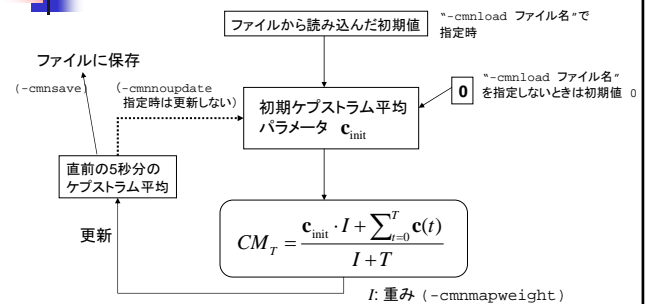
バージョン	Fast-PTM	Standard-Tri
3.4.2	2.01	7.05
3.5.2	2.04	7.02
3.5.3	1.76	4.56

JNAS200文の平均実行時間(秒)、Pentium4 3GHz

②MAP-CMN (3.5.1)

- マイク入力時のCMNを改善
- 従来法
 - 直前の5秒間のケプストラム平均
⇒ 次の発話のケプストラム平均として適用
 - 話者交代時にミスマッチ
- MAP-CMN
 - 初期ケプストラム平均パラメータを用意
 - 初期フレームは上記のパラメータでCMN
 - 入力が伸びるにつれて発話自身のCMNに近づける
 - 直前の発話で初期パラメータを更新

アルゴリズムとオプション



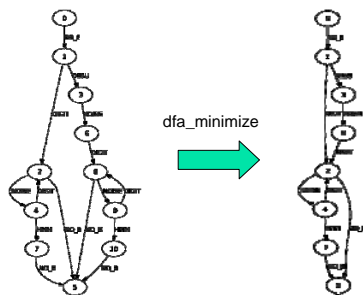
③グラフ出力の改善 (3.5.2)

- アルゴリズムの改善
 - 仮説マージ時にスコアが高いパスを優先するよう修正
 - グラフの深さによるカットオフ (-graphcut 深さ)
 - 後処理の境界確定ループを一定回で打ち切り (-graphboundloop 回数)
 - 単語間トライフォンを考慮したグラフ生成 (-graphrange -1)
- デフォルト
 - -graphcut 80 -graphboundloop 20

④認識用文法の最適化: dfa_minimize

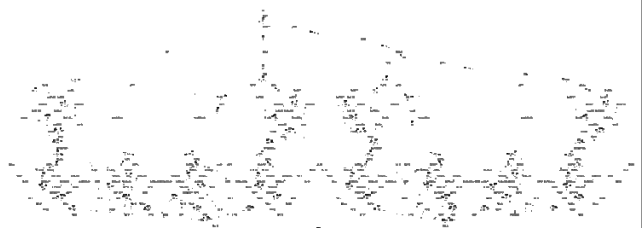
- 文法の有限オートマトン (DFA) を最小化
 - 現在の mkdfa.pl は冗長な状態を出力
 - 最小化により冗長な状態を縮退
 - 認識処理の効率の改善
- 使い方
 - A) 既存の DFA を最小化
`dfa_minimize infile.dfa -o outfile.dfa`
 - B) 3.5.3以降の mkdfa では自動的に実行される

サンプル文法 digit



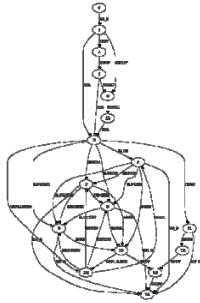
サンプル文法 price

87 nodes, 267 arcs



最小化後

17 nodes, 39 arcs



⑤メモリの最適化・遅延改善

- **メモリの最適化**
 - bmalloc() によるメモリのブロック確保
 - 音響モデルのメモリのカプセル化
 - 第2パスで捨てられた仮説のメモリを再利用
- **遅延改善**
 - サイクルバッファによる1次差分・2次差分計算
 - DirectSound の使用 (Windows)

⑥MFCC特徴量のフルサポート

- MFCC の全抽出パラメータを指定可能に
 - HTK との互換性を確保
- MFCC ベースの音響モデルであれば基本的に Julius で音声の直接認識が可能
- ただしパラメータのデフォルト値が Julius と HTK で異なる点に注意

HTK Config との対応表(1)

HTK Option	Description	HTK default	Julius default	How to set
TARGETKIND	Parameter kind	ANON	MFCC	Only MFCC, qualifiers (_E, _N, etc.) automatically set from AM header
NUMCEPS	Number of cepstral parameters	12	-	automatically set from AM header
SOURCERATE	Sample rate of source waveform in 100ns units	0.0	625	"-smpPeriod value"
TARGETRATE	Sample rate of target vector (= window shift) in 100ns units	0.0	160	"-fshift samples" (*)
WINDOWSIZE	Analysis window size in 100ns units	256000.0	400	"-fsize samples" (*)
ZMEANSOURCE	Zero mean source waveform before analysis (frame-wise)	F	F	"-zmeanframe" to enable, "-nozmeanframe" to disable.
PREEMCOEF	Set pre-emphasis coefficient	0.97	0.97	"-preemph value"
USEHAMMING	Use a Hamming window	T	T	Fixed

HTK Config との対応表(2)

HTK Option	Description	HTK default	Julius default	How to set
NUMCHANS	Number of filterbank channels	20	24	"-fbank value"
CEPLIFTER	Cepstral liftering coefficient	22	22	"-ceplif value"
DELTAWINDOW	Delta window size in frame	2	2	"-delwin value"
ACCWINDOW	Acceleration window size in frame	2	2	"-accwin value"
LOFREQ	Low frequency cut-off in fbank analysis	-1.0	-1.0	"-lofreq value", or -1 to disable
HIFREQ	High frequency cut-off in fbank analysis	-1.0	-1.0	"-hifreq value", or -1 to disable
RAWENERGY	Use raw energy	T	F	"-rawe" / "-norawe"
ENORMALISE	Normalise log energy	T	F	"-enormal" / "-noenormal" (**)
ESCALE	Scale log energy	0.1	1.0	"-escale value"
SILFLOOR	Energy silence floor in Dbs	50.0	50.0	"-silfloor value"

⑦HTK Config の読み込み

- HTK Config ファイルを直接読み取り可能
 - 音響モデル学習時の設定ファイルをそのまま Julius に与えれば良い

```

julisus ... -htkconf ConfigFile
            
```
- バイナリHMMのヘッダに埋め込み可能
 - 特徴量条件を埋め込んでバイナリHMMを生成
 - 認識実行時は何も指定する必要がない

```

mkbingram ... -htkconf ConfigFile
            
```

⑧HTKからJulian への文法変換

- ツール "slf2dfa" を公開
 - HTK の文法・辞書をJulian形式へ自動変換
 - Julianユーザ: 正規表現による簡潔な文法記述が使える
 - HTKユーザ: Julian への移行が簡単
- Julius のホームページよりダウンロード可能

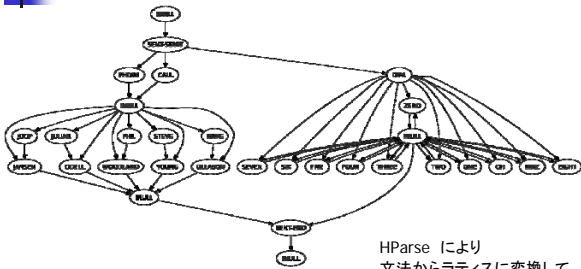
HTKの文法表現

正規表現

```
$digit = ONE | TWO | THREE | FOUR | FIVE |
        SIX | SEVEN | EIGHT | NINE | OH | ZERO;
$name = [ JOOP ] JANSEN |
        [ JULIAN ] ODELL |
        [ DAVE ] OLLASON |
        [ PHIL ] WOODLAND |
        [ STEVE ] YOUNG;
(SENT-START (DIAL <$digit> | (PHONE|CALL) $name) SENT-END)
```

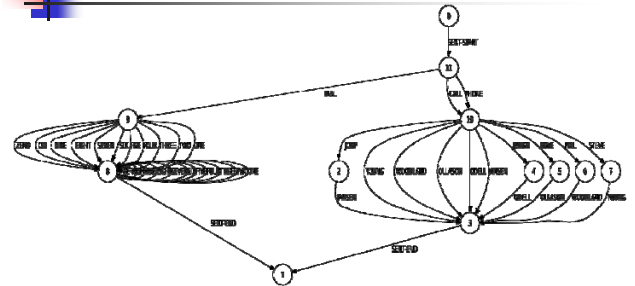
音声ダイヤルタスク用認識文法(番号もしくは人名)

Standard Lattice Format



HParse により
文法からラティスに変換して
使用する

SLFからDFAへ変換



(実際は逆向き)

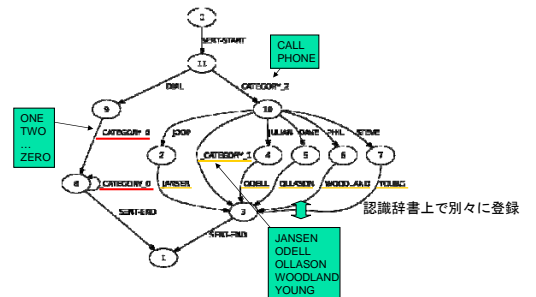
HTK SLF と Julian DFA の相違点

1. Moore型
2. NULL遷移可能
3. Left-to-right
4. 辞書から文法にある単語のみ読み込んで認識
5. 単語単位の制約



1. Mealy型
2. NULL遷移不可
3. Right-to-left
4. 辞書の単語を全部読み込んで認識
5. カテゴリ単位の制約
制約が同じ単語集合をよりコンパクトに扱える

単語カテゴリの自動検出



(本来は逆向き)

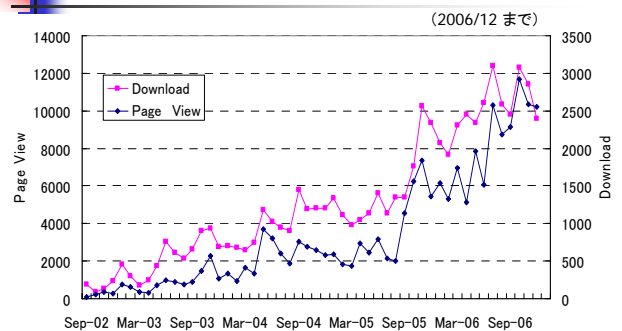
⑨ ホームページの再構成



<http://julius.sourceforge.jp/>

25

月別ダウンロード数



26

バージョン履歴

- 3.5.3 (2006/12/29)
 - 高速化(割り算排除)・メモリ最適化
 - 文法最小化ツール・slf2dfa
 - HTK Config ファイル対応、音響モデルへの埋め込み
 - 単語間トライフォン依存グラフ
- 3.5.2 (2006/07/31)
 - DirectSound対応、グラフ生成安定化
 - 第1パスグラフ生成
- 3.5.1 (2006/03/31)
 - MFCC HTK互換性フルサポート、MAP-CMN
デルタ計算遅延改善

英語版音声認識キット

- Julius/Julian
- Linux/Windows

英語版音声認識キット

- WSJコーパスでモデル学習
 - 音響モデル...EDANZ, 7000x16
 - Si284set (37K発話, 62時間)
 - 単語辞書...39(CMU phoneset)+sp+sil
 - ディクテーション用言語モデル...20K trigram
- Juliusの設定
 - マルチパス音響モデル
 - ショートポーズの処理

WSJ評価結果

- ARPA Hub2, Nov. 1993
 - 5K: 215発話、7.3秒
 - 20K: 213発話、7.0秒

	Julius	HTKデコーダ
5K	90.5 /4RT 87.4 /RT	90.8 /10RT 84.4 /RT
20K	82.1 /4RT 79.3 /RT	NA

Julian用サンプル文法(英語版)

- digit: 連続数字
- number: 数字
- date: 日付
- time: 時間
- persons: 人数
- price: 値段
- yesno: Yes/No
- spell: 音素タイプ
- attendant: 受付
- fruit: 果物注文

用語の説明

- Julius: 大語彙連続音声認識のフリーソフトウェア
 - 当初はディクテーション向けのN-gram言語モデル対応のものだったが、現在は下記Julianも統合
- Julian:
 - (人手による)記述文法対応の認識プログラム
- SAPI: Speech API
 - マイクロソフト社策定のAPI、Windows XPに標準搭載
- SALT: Speech Application Language Tags
 - マイクロソフト社などが策定している、HTMLブラウザで音声認識・合成を行うためのタグ
 - Speech Application SDKに含まれる
- SRM: Speech Recognition Module
 - JuliusのGalatea Toolkitのためのインタフェース

関連Webページ

- Julius
 - 最新版(SAPI版含む)フリーダウンロード
 - <http://julius.sourceforge.jp/>
- 連続音声認識コンソーシアム(CSRC)
 - 最終版を一般頒布中(有償)
 - <http://www.lang.astem.or.jp/CSRC/>
- 日本語ディクテーションツールキット
 - 最終版は「音声認識システム」(オーム社)の付録CD-ROM
 - <http://www.ar.media.kyoto-u.ac.jp/dictation/>
- リーディングプロジェクト e-Society基盤ソフトウェア
 - <http://cif.iis.u-tokyo.ac.jp/e-society/>
- Microsoft Speech Application SDK
 - SALTIには必要
 - <http://www.microsoft.com/speech/>