やってみよう Linux で音声認識
~ オープンソースな認識エンジン Julius ~

李 晃伸 (京都大学)
ri@kuis.kyoto-u.ac.jp

1999 年 12 月 18 日 (土)

1 Julius とは

Julius とは , 数千から数万語の認識が可能なフリーの大語彙連続音声認識プログラムです . これは日本語の音声認識の研究やソフトウェア開発を支援するツールキット作成プロジェクト「日本語ディクテーション基本ソフトウェア」[注1] の成果物です . このプロジェクトには奈良先端大 , 名古屋大 , 京都大 , 電総研 , そのほか企業の研究所を含む様々な音声に関する研究機関が参加しており , 3 年前から共同で取り組んでいます . これらの成果物は全てフリーで入手できます .

現在の認識性能は 90% 以上と商用のソフトに引けをとりません . メインの動作環境は Linux で現在も開発が進められており , 簡単な SDK を含め来年の 3 月に正式公開が予定されています . また現時点でのベータ版キット (1998 年度版) とお試し版が Web 上から入手できます .

今回の発表では , Julius の開発者である筆者が , 音声認識関連の話題とともに , このプロジェクト全体と Julius , Linux との関わりについて紹介します .

2 背景

人間の発声をコンピュータに理解させる音声認識の研究は , かなり昔から行われています . 特に最近になって , 音声を使った商品が幾つか世に出回りはじめ , 一般から注目を浴びるようになってきました . 現在世に出ている音声認識ソフトウェアは以下のようなものがあります .

- ViaVoice (IBM)
- SmartVoice (NEC)
- Dragon NaturallySpeaking (Dragon Systems)

これらは基本的にディクテーション , つまりいわゆる「音声ワープロ」を目指しており , 数万語の語彙を対象としてとにかく日常の発声を全て文字に書き落す「大語彙連続音声認識」の分野です . Julius もこのグループに含まれます .

そのほかエンターテイメントの分野では

- ピカチュウげんきでチュウ (Nintendo64)
- シーマン (DreamCast)

などの製品が発売されています . こちらは小規模で発声も基本的に単語単位です . 語彙はせいぜい数百程度ですが , 対話の流れなどに重点がおかれています .

このように近年注目されるようになった技術的な背景としては , 競争によってここ数年で音声認識の各要素技術が急激に進歩したことが挙げられます .

- 大量のデータベースを用いた統計的モデル化のアプローチの成功
- 高性能な認識アルゴリズムの実現と精錬
- 計算機の性能向上

しかし技術としてはまだまだ枯れてはおらず , 今後も様々な応用が考えられるチャレンジングな分野だと思います .

Linux では現在 , 音声認識関係のアプリは大変少ないです . 英語では音声合成や認識のソフトウェアがいくつかありますが[注2] , 特に日本語に関してはほとんど無いといってよいでしょう . Linux で日本語が認識できるフリーの認識ソフトがあれば , 下記のような様々な応用が考えられますし , いろいろ幸せになれる人も多いかもしれません : -)

[注1] http://www.itakura.nuee.nagoya-u.ac.jp/~takeda/IPA/

[注2] IBM が ViaVoice SDK for Linux のベータ版 (英語のみ) を配布しています:
http://www-4.ibm.com/software/speech/dev/sdk_linux.html

- 音声操作デスクトップ環境
- Linux Box の音声遠隔操作
- 電話の自動応答サービスのサーバ

3 大語彙連続音声認識のしくみ

ここで，まず一般的な大語彙音声認識システムの構成について少し説明します．音声認識システムは一般に，図 1 にように言語モデル・音響モデル・解探索アルゴリズムの 3 つのモジュールから構成されます．

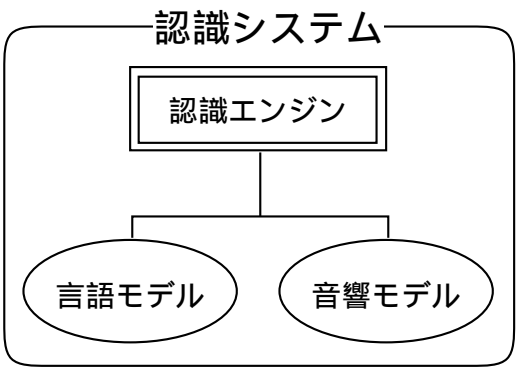

図 1: 一般的な認識システムの構成

言語モデルは単語列の言語的な出現確率を与えます．例えば「今日」の後ろには，「は」が出る確率が 0.6,「で」が出る確率は 0.2... といったようにです．これらの確率は，名詞，動詞，助詞などの文法的な構文構造から人手で与えることも出来ますが，最近は新聞記事などの大量のデータベースから形態素解析を行って，単純な単語列の出現確率を統計的に推定する方法が主流です．また音響モデルは，各単語について，入力された音声に対してどの程度マッチするかを確率で与えます．これらの言語モデルと音響モデルの確率を掛け合わせたときに，最も可能性の高い単語列が認識結果となります．

この最尤単語列を見つけ出すのが，認識エンジンである Julius の役割です．登場し得る候補の数は非常に膨大です．入力が N 単語からなるとすると単語列の候補は基本的に (語彙数の N 乗) 個存在するわけで，それらをいちいち入力とマッチングしていたら 1 文の入力に 30 分かかるということにもなりかねません．如何に効率よく最も良い単語列を見つけ出すかが認識エンジンのアルゴリズムのカギとなります．

なお，日本で「音声ワープロ」というと，音声をいったんひらがなに認識してから漢字に変換するという「音声→ひらがな→漢字かな混じり」の流れを想像される場合が多いようです．しかし Julius を含め多くの認識エンジンは，音節ごとではなく単語単位で，各単語に対応する音の読みを直接使って認識する「音声→漢字かな混じり」の方式をとっています．これは言語的な情報を全く使わずに任意の音節の並びとして音声を認識するのは難しいためで，高い精度を達成するには単語の語彙の制限やその単語間の接続制約を早い段階で最大限利用するほうがよいからです．

4 「日本語ディクテーション基本ソフトウェア」開発プロジェクト

Julius は，「日本語ディクテーション基本ソフトウェア」作成プロジェクトの成果物です．このプロジェクトの目的は，日本語の大語彙連続音声認識のためのオープンソースなツールキットを作成することです[1]．

高精度な言語モデルと音響モデルを作るには，大量のデータベースが必要です．現在は音声データを CD-ROM 数十枚分，テキストデータを数 GB 分使ってモデルを作成していますが，まだまだ十分ではないとされています．また素早く認識を行うには，認識アルゴリズムの工夫とその高度なプログラム実装が必要です．

これらの作成を研究機関ごとに行うのは多大な労力と技術が必要ですし，無駄ともいえます．そこで各モジュールごとに分担して共有可能な標準モジュールを作ることで，個々の作成の負担を軽減することができます．また，個々で作ったモデルや認識エンジンを入れ替えることで既存のものとの相互比較が簡単に行えるようになります．またそのような標準的な環境を無償で一般に提供することで，新たな音声アプリケーションの開発を促す効果もあります．

このプロジェクトは 3 年前から 情報処理振興事業協会 (IPA) の支援のもとで，奈良先端大，名古屋大，京都大，電総研，そのほか企業の研究所を含む様々な音声研究の機関が共同で参加しています．2000 年 3 月に終了の予定ですが，その後もなんらかの形で活動を続けていくことが計画されています．終了時には最終成果物としてこれらのモデルや Julius，および認識システム構築のための周辺ツールがまとめて正式リリースされる予定です．

トータルの性能については，モデル・エンジンともに最先端の研究成果を常に盛り込みつつ精力的な改善が続けられており，商用のものにも負けない高精度なレベルにまで

達しつつあるといえます[2].

4.1 ライセンスについて

このツールキットのライセンス文書は

http://winnie.kuis.kyoto-u.ac.jp/pub/julius/toolkit_license.html

にあります．以下に私の解釈による重要部分を載せますが，正確なところは必ず上記の文書をあたってください．

1. 商用利用 OK．一部を組み込んだアプリを作成し配布・販売することに関しても著作権を行使しない．
2. ただし著作権表示の全てを配布物に添付すること
3. 無保証，自己責任での運用
4. 生命，身体，財産等に重大な損害が発生する危険を有するシステムに使用しないこと．

なおキット内に納められている各ツール (Chasen 等) については独自の license を有するものがあります．

4.2 キットの入手方法

現在 1998 年度時点での成果をベータ版として無償配布しています．これには以下のものが含まれています．

- Julius ver.2.1 ソースアーカイブ
- 様々なタイプの言語モデルと音響モデル
- 形態素解析・読み付与ツール (Chasen, Chawan など)
- その他評価用ツールなど
- ドキュメント

現在キットの入手には Web 上で登録をしていただくようになっています．また登録することで，将来にわたるバグ情報やアップデート情報を受け取ることができます．また実際の配布は，CD-ROM の郵送という形になっています．

5 Julius で音声認識をやってみる

この「日本語ディクテーション基本ソフトウェア」は基本的に音声関連の研究者やアプリケーション開発者向けのキットで，容量もかなり大きいです．そこでとりあえず気軽に Linux で音声認識を試してみることができるように，1998 年度版キットから最低限のものを抜き出して「お試し版」を作成しました．まだまだ未完成ですが，興味のある方はぜひ試してみてください．

5.1 お試し版ダウンロード先

以下の URL から辿れます (右端で折り返しています).

http://winnie.kyoto-u.ac.jp/pub/julius/trial-kit-v1998/index.html

5.2 動作条件

動作に必要なスペックです：

CPU Pentium 300MHz 以上，400MHz 以上を推奨
メモリ 64MB 以上必須，128MB 以上を推奨
サウンドカード **16bit 16kHz 録音** 可能なもの

サウンドドライバは OSS コンパチブルなドライバ(カーネルドライバおよび OSS/Linux) もしくは ALSA が利用できます．16bit, 16kHz でマイク録音できることが必須です．いわゆる「SoundBlaster Pro 互換」の設定ではうまく動きません．

快適な速度で動作させるにはかなりの高スペックを要します．目安としては，Pentium 400MHz 以上, メモリ 128MB 以上ならしゃべり終わるとほぼ同時に認識が終了するでしょう．

5.3 注意点

認識はノイズや音割れ，音の歪みなどに非常に敏感です．お試し版の公称の認識精度は約 85% ですが，発声の仕方や周囲の環境，サウンドチップやマイクの良し悪し，そして個人差などによって大きく変動します．状況によっては全く認識できないこともあります．これについては今後の課題です．

なお Julius に関するお問い合わせは julius@kuis.kyoto-u.ac.jp で受け付けています．このお試しキットや Linux に関する動作報告・バグ報告・ご意見・ご感想・提案などは，直接私こと 李 (ri@kuis.kyoto-u.ac.jp) 宛にお願いします．動作しない場合はもちろん，うまく動いた場合でも，動作環境を上記までメールしていただけると嬉しいです．

6 現状の問題点

実用的な認識システムとして足りない機能：

- 環境適応や話者適応 (エンロール)

- 辞書登録や言語モデルのカスタマイズ (専門知識を要する)

認識エンジンの今後の課題：

- 認識精度
- 認識速度
- ドキュメントの整備

7 API 整備とオープンソースについて

これまで ,プロジェクト全体の指針やソース自身は「フリー」なのですが ,Julius の開発はクローズドな体制 ,というより著者一人で行ってきました . これは当初 Julius のコアである認識アルゴリズムがまだまだ発展途上で ,著者自身の大学での研究を兼ねて実装されており [3], 複数人数で分担する体制になかったためです . 今でも Julius の設計・実装・評価・バグ修正・メンテナンスなど ,基本的には一人で行っています .

最近になって認識エンジンのコア部分の仕様がほぼ固まってきたので ,次のステップとしては ,様々なアプリケーションへの組み込みを考えて API を整備しライブラリ化するべきだと考えています . 音声認識を組み込んだアプリが誰でも簡単に作れるようになれば ,いろんな人が音声認識に触れられるようになるし ,おもしろい応用が生まれたり研究上の新しい知見が得られるかもしれません .

そこで ,今後は Julius の API の策定や機能強化など ,オープンソースな開発体制で進められたら ,と考えています . どれほどの規模でどれだけのことができるかは分かりませんが ,興味あるみなさんの協力が得られればと考えています .

8 まとめ

Linux で動作する大語彙連続音声認識のエンジンである Julius と IPA の日本語ディクテーション基本ソフトウェアの取り組みについて紹介しました . これらの現在の成果は全てソース付きでフリーで入手することができます . 今後もこのプロジェクトは継続する予定です . また Julius 開発の今後について ,興味のある方は是非著者まで連絡いただけたらと思います .

最後になりましたが ,著者個人の Julius を作る原動力はやはり「やってて楽しい」です . この音声認識の楽しさをみなさんにも感じていただけたらと思う次第です .

謝辞

日頃ご指導いただく京都大学大学院情報学研究科の河原助教授に深く感謝致します . また Julius の開発は情報処理振興事業協会 (IPA) の「日本語ディクテーションの基本ソフトウェアの開発」(代表者：鹿野清宏 奈良先端科学技術大学院大学教授) の援助を受けて行われました . 日頃ご助言 ,ご討論いただく関係各位に感謝致します . そしてこの開発環境をフリーで提供していただいている Linux および GNU ツール群に感謝致します .

参考文献

[1] 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 伊藤克亘, 伊藤彰則, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏. 日本語ディクテーション基本ソフトウェア (97 年度版). 日本音響学会誌, Vol. 55, No. 3, pp. 175–180, 1999.

[2] 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 伊藤克亘, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏. 日本語ディクテーション基本ソフトウェア (98 年度版) の性能評価. 情報処理学会研究報告, 99-SLP-26-6, 1999.

[3] 李晃伸, 河原達也, 堂下修司. 単語トレリスインデックスを用いた段階的探索による大語彙連続音声認識. 電子情報通信学会論文誌, Vol. J82-D-II No.1, pp. 1–9, 1999.

著者略歴

李 晃伸：Julius の認識アルゴリズム担当 , メインプログラマー . Linux 歴 5 年 . 音声研究歴はもう 5 年 . 初めての Linux インストールは ThinkPad 230Cs にフロッピーディスク経由だった . 今は Debian 好き . 京都大学大学院情報学研究科博士課程在籍 . 京都生まれの兵庫育ち , 26 歳 .