

音声入力における対話的候補選択手法

遠藤 拓 寺田 実

東京大学大学院 情報理工学系研究科 知能機械情報学専攻

1 はじめに

音声入力はキーボード入力に比べて疲労が少なく入力速度も速いので長い文章の入力には適している。ところがディクテーションソフトを日常的に文書作成に利用する人は少ない。その原因の1つとして、我々は音声認識で生じる誤認識を考えた。キーボードを使った入力に比べ、音声認識の誤認識はユーザが予期せぬものであり、その修正には時間がかかって面倒である。この修正にかかる時間を短縮することでユーザの負担を軽減することが本研究の目的である。

そのために本研究では、認識結果の単語に対する候補単語をリスト表示する機能について考える。従来のディクテーションでは、認識エンジンが内部的に生成する複数の候補のうち、最良の一つを除いて残りをすべて捨てていたが、これを捨てずに活用して、(1) 誤認識を修正する際の単語候補として利用し、(2) その単語を含むような文全体の認識結果を新たな結果とする。前者によって、誤認識単語の修正候補として正解を提示できる可能性が高まり、後者によって、一つの単語のみを修正することによって発生する、日本語としての不自然さを解消できると考える。

2 認識候補のリスト表示

2.1 認識候補

ディクテーションに用いられる大語彙連続音声認識では、認識過程でいくつもの仮説が立てられる。認識エンジンは音響モデルと言語モデルを基にしてそれらの仮説中で確率的に最も尤もらしいものを認識結果として残し、それ以外の仮説は棄却する。もし認識エンジンが出力した認識結果に誤りが含まれている場合、棄却された仮説の中に正解(ユーザが意図した単語)が含まれていることが多い。[1]で我々が実装したシステムでは、誤認識単語の半数以上に対する正解が認識エンジン内の仮

説に含まれていた。そこで、認識エンジン内で棄却された仮説中の単語のうち、始点と終点の時間が誤認識された単語と近いものをその単語の認識候補として表示することにした。

2.2 一単語のみを置換する場合の問題点

ある誤り単語を単純に別な認識候補に置き換える機能のみを実装した場合、以下の2つの問題点がある。

第一点は、キーボード入力でのかな漢字変換のように単語の境目を変更することができないため、誤り単語とユーザが希望する単語の長さが異なる場合には修正することができない点である。

例えば「富士山麓にオウム鳴く」と入力して、

| 富士山麓 | に | 大村 | 君 |

と認識されたとする。「大村」と「オウム」では単語の長さが異なるので認識候補に「オウム」が含まれる可能性は低い。仮に候補に「オウム」があり、それを選んだとしても、

| 富士山麓 | に | オウム | 君 |

となる。次に「君」を変更するわけだが、「君」の候補で「鳴く」が出てくる可能性はこれまた低いと考えられる。

第二点は、時折、修正後の新たな結果に不自然な点が生じることである。

例えば「マイクロバスに乗る」と入力したものが、

| マイク | と | バス | に | 乗る |

と認識された場合、「マイク」を正しい「マイクロ」に置換すると文章が、

| マイクロ | と | バス | に | 乗る |

となってしまう。この場合、単語の置換と同時に後ろにある助詞も消える方がよい。また、音声認識ではユーザの思いもよらない突拍子もない結果が現れることがある。そのような場合に、誤り始めの一単語を正しいものに変えることで文全体も正しく置き換わると大変便利である。

以上の問題を解決するために我々は、認識候補をリスト表示する機能に以下の2つの機能を追加することにした。

⁰Candidate selecting interface for speech input

Taku Endo, Minoru Terada

Department of Mechano-Informatics, School of Information Science and Technology, University of Tokyo

2.3 連鎖的置換機能

ユーザがある誤認識単語に対する認識候補を選択すると、その単語だけでなく後に続く残りの文も連鎖的に（言語モデル的に）最適なものに置き換わる機能である。これを連鎖的置換機能と呼ぶことにする。仮にある単語を正しいものに置き換えたことで、元は正しかった後ろの文に誤りが生じた場合には、新たな誤りも修正しなければならなくなる。修正の手間が増えることも考えられるが、これは“誤認識を修正した後の文の不自然さを修正する”手間とのトレードオフだと考えている。

2.4 単語境界の変更機能

日本語入力のかな漢字変換では単語の区切りを変更できる。これと同様の機能を音声認識の他候補を選ぶ際にも利用できるようにした。この機能と2.3の連鎖的置換機能と組み合わせれば、2.1の例「富士山麓に大村君」で、「大村」の右端を左に移動させて「オウム」にした直後、「オウム」の後ろも自動的に「鳴く」に変わるということも可能になる。

3 実装

我々は、音声認識エンジン Julius[2] を使って前述の機能を実装した [1]。しかし [1] での実装では認識エンジンの出力単語に対して ID となる番号を付け、それに対してコマンドラインからキーボード入力での単語の指定や「候補表示」、「境界変更」といった操作を行っていた。今回新たにマウスを用いてこれらの操作を行なうための GUI を実装した。

音声入力した結果を単語毎に“|”で区切って表示し、ユーザが誤認識単語をマウスクリックすると、その下に認識候補がリスト表示される。ユーザがその中から望む単語を選択すると、元の単語と選択された単語が置換される。その際、連鎖的置換機能によりその単語より後ろにある単語も、認識エンジンが尤もらしいと判断する単語列に置き換わる。ユーザが単語をクリックして候補をリスト表示した後、その単語の両端にある“|”のどちらかを左右どちらかにドラッグすると、その単語の境界を変更した際の候補となる単語がリスト表示される。

4 評価

現時点ではアプリケーションを作成したものの、まだ被験者を募っての使用実験は行っていない。今

「富士山麓にオウム鳴く」と入力したが
「富士山麓に大村君」と出力された場合

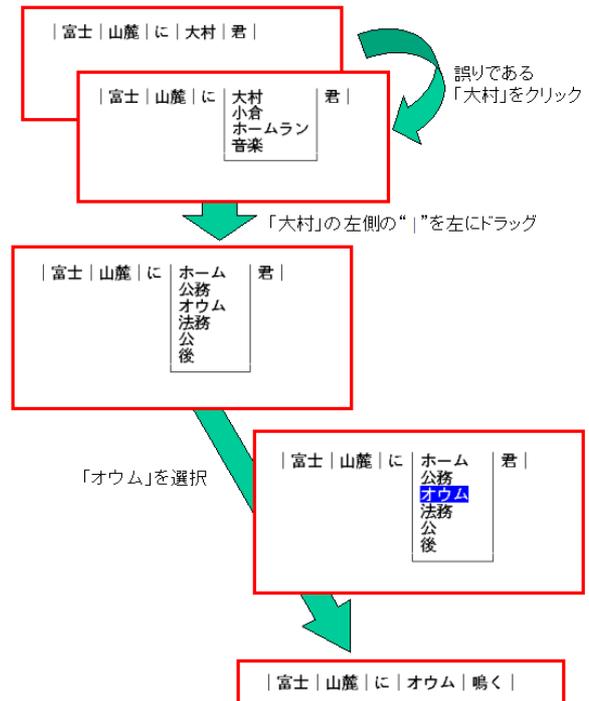


図 1: GUI 操作のイメージ

後はユーザに実際に本機能を使用してもらった上でのアンケートによる主観的な評価と、コーパスを用いて、誤認識が生じた際に候補中に正解が含まれる確率・本システムの機能が有効となる割合等の定量的評価を計画している。

5 結論

音声認識の誤認識修正の手間を軽減するために認識候補をリスト表示するという機能を取り上げ、利便性向上のために新たに連鎖的置換機能と単語境界の変更機能という2つ機能を提案し、実装してGUIをつけた。ユーザによる評価実験はこれからだが、筆者の使用した感想では本機能が誤認識の修正に役立つと考えている。

参考文献

- [1] 遠藤 拓, 寺田 実. 音声認識における他候補表示についての検討. 第10回インタラクティブシステムとソフトウェアに関するワークショップ. pp. 145-146. 日本ソフトウェア科学会. 2002.
- [2] 大語彙連続音声認識デコーダ Julius, <http://winnie.kuis.kyoto-u.ac.jp./pub/julius/>