

接続情報を利用した多段階分割復元手法の 話題変化への追従性

浦南至徳[†] 荒木健治^{††}
北海道大学大学院工学研究科[†]

宮永喜一[†] 枋内香次[†]
北海学園大学工学部^{††}

1 はじめに

計算機への日本語音声入力を考える場合、音声認識結果である誤りを含んだ音素列から、言語情報を利用して訂正を加え、漢字かな混じり文へ復元する処理が重要となってくる。このような復元処理を行なう手法として、我々は接続情報を用いた多段階分割復元手法を提案し [1][2]、実際の連続音声認識結果への応用を行なってきた。また、音声認識率と本手法の復元能力の関係についてシミュレーション実験を行ない、その結果、学習と復元処理に類似性の高い文章を選んだ場合において、良好な復元結果を得るためには80%以上の音声認識率が必要であることを確認した。

しかし、ワードプロセッサへの日本語音声入力などを考えると、使用分野を限定した評価だけでは不十分であり、分野を限定しない連続音声認識・理解を考えた場合の話題変化に対し追従できるかどうか重要となる。

本稿では、学習機能を持った本手法の復元能力を、話題変化すなわち対象分野の変化に対する追従性という点で調べた結果について述べる。具体的には、2つの分野に属する論文、計12編に対し、実際の音声認識結果の誤り傾向に基いた誤りを付加し、本手法により復元を行ない評価を行なった。

2 音素列シミュレーション作成

性能評価実験には多量の音素列が必要であるが、十分な音声データが手元になく、実際の音声認識により音素列を準備することができなかったため、少量の音声認識結果に基づく音素列作成シミュレーションにより誤りを含む音素列データを作成する。

2.1 誤り傾向の獲得

実際の音声認識結果には、認識アルゴリズムなどにより誤り方に、ある傾向が存在すると考えられる。よって、元音素列から誤りを含む音素列を作成する際、ランダムに誤らせるのでは問題がある。そこで、本稿

ではO(n)Viterbi法によるHMM音声認識装置により実際に音声認識を行ない、その結果と元音素列を比較することで誤り方の傾向を獲得し、誤り傾向辞書へその対応と誤り確率を記述しておき、音素列作成の際に利用する。その際、付加、脱落誤りにも対応できるように誤りと判断した音節の前後の音節も同時に切り出してある。誤り傾向の抽出法、獲得される結果の一例をFig.1に示す。

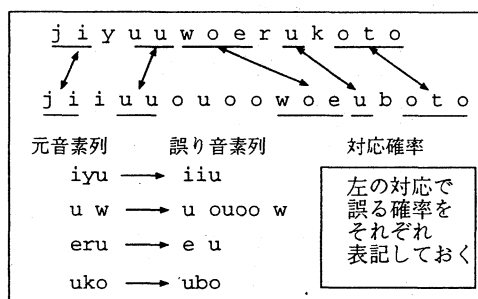


Fig.1 誤り傾向抽出法、誤り傾向辞書

2.2 シミュレーション

2.1節で述べた誤り傾向辞書を用いて、シミュレーションを行なう。また、この作業を行なうことで実際の音声認識結果と類似した結果が得られると考える。以下、簡単に処理手順を示す。

1. 元音素列データに対し音節毎にセグメンテーションを行なう。
2. 各音節とその前後の1音節ずつを連結させたものが誤り傾向辞書中に存在するか調べる。
3. 辞書中に存在する場合は、対応する誤り確率で誤らせるかどうか決定する。
存在しない場合は誤り確率0%とする。
4. 全ての音節に対し2,3を繰り返す。

3 復元手法

本手法では接続情報を用い、確実性が低いと考えられる音素数の少ない語を連結し、比較的高い確実性をもつ語として復元することが可能であり、多段階に確実性の高い語から当てはめを順次行なっていくことで、その後の処理において、候補数を抑制し復元効率を向上させることが可能になる。また、復元結果と原文とを比較し、学習処理を行ない、未登録の単語、単語情報の登録、既登録の単語情報の更新を行なう。よって、本手法は対象とする分野が変化しても、適宜、適応していくものと考えられる。復元処理の流れを Fig.2 に示す。

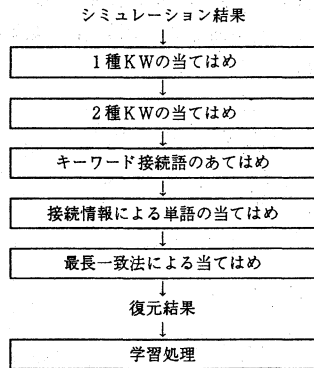


Fig.2 復元処理の流れ

4 実験と結果

4.1 予備実験 1 音声認識と誤り傾向の獲得

1. 音声認識

使用ツール

日本情報処理開発協会発行研究用連続音声データベース利用支援ツール内の O(n)Viterbi 法による連続音声認識ツール

使用データ

ATR 音声データベース文セット

男性話者 10KHz 200 文

結果

音声認識率：約 70 %

2. 誤り傾向の獲得

2.1 節で説明した方法で、音声認識結果と正しい音素列の比較を行ない、誤り傾向を獲得し、誤り傾向辞書を作成した。

4.2 予備実験 2 シミュレーション

2.2 節で説明した方法で、正しい音素列に誤りを付加した。また、予備実験 4.1 において実際の音声認識結果の認識率は約 70 %であったので、シミュレーションは、設定認識率 70 %として行なった。以下に使用データを示す。またシミュレーション結果の認識率を表 1.1、表 1.2 に記す。

使用データ

情報処理工学関連（以降、分野 A）の論文 6 編
機械工学関連（以降、分野 B）の論文 6 編

結果

A1	A2	A3	A4	A5	A6
71%	70%	70%	72%	71%	72%

表 1.1. 情報処理関連分野の各認識率

B1	B2	B3	B4	B5	B6
71%	72%	71%	72%	72%	71%

表 1.2. 機械工学関連分野の各認識率

4.3 復元実験

あらかじめ分野 A のデータ（復元実験に使用するものとは異なる）で、十分に学習を行なっておく。ついで、4.2 節で作成した誤りを含む音素列を、3 節で説明した復元システムへの入力として、復元実験を行なった。復元の際は、A1、A2、…、A6、B1、B2、…、B6 の順番で復元を行なうこととした。

4.4 実験結果

復元結果は正当ではめ率と復元精度をもって評価を行なう。結果の一部を Fig.3.1.1、Fig.3.1.2、Fig.3.2.1、Fig.3.2.2 に示す。Fig.3.1.1 と Fig.3.1.2 は正当ではめ率について、Fig.3.2.1、Fig.3.2.2 は復元精度についての結果である。各図の横軸は復元したデータの行数である。ここで、復元精度と正当ではめ率は次式で定義する。

$$\text{正当ではめ率} = \frac{\text{正しく当てはめられた音素数}}{\text{総音素数}}$$

$$\text{復元精度} = \frac{\text{正しく当てはめられた音素数}}{\text{当てはめに用いられた音素数}}$$

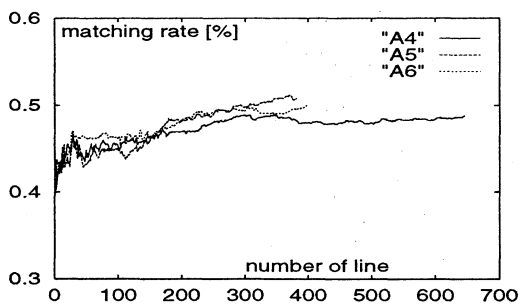


Fig.3.1.1 正当ではめ率 (A分野)

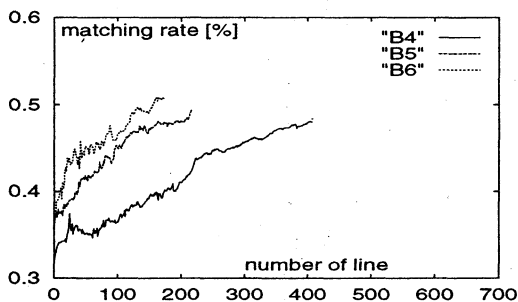


Fig.3.1.2 正当ではめ率 (B分野)

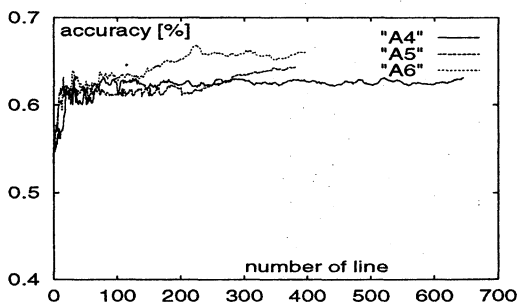


Fig.3.2.1 復元精度 (A分野)

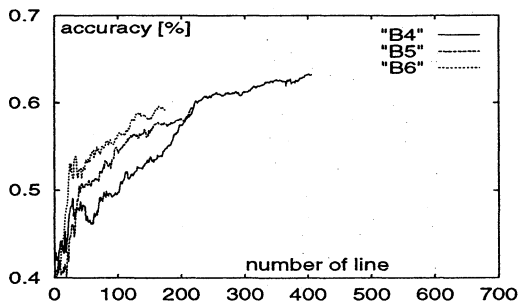


Fig.3.2.2 復元精度 (B分野)

4.5 考察

復元精度、正当ではめ率の結果共に、話題変化への追従性がよく表れていると思われる。すなわち、事前に十分に学習されてある分野のデータに対する結果は、グラフの変化が小さく、学習されていない分野のデータに対しては、最初は低い復元精度しか得られないが、復元が進むにつれて学習が進み、復元精度が向上している。後半部分では、学習済み分野のデータに対する結果よりは若干低いものの、それに近い復元精度まで上昇していることがわかる。

また、B分野の各文章の冒頭部分の結果が非常に低くなる原因は、キーワードのような使用頻度の高い単語は、各文章固有のものであることが多いため、文章が変わると、再度、その文章のキーワードを決定するまでの間、確実性が落ちるためであると考えられる。この現象がA分野において小さいのは、十分な学習を前もって行なうことで、A分野において、全体として使用頻度が高い語が、キーワードとして決定されていたためであると考えられる。

5 おわりに

音声認識結果に類似した誤りを含む音素列データを用意し、そのデータを用いて、我々の提案している、接続情報を用いた多段階分割復元手法の話題変化への追従性について実験的に調べた。

その結果、話題が大きく変化すると、一旦、復元精度は落ちるが、さらに復元が進むにつれて精度が上がり、話題変化に適応していけることが明らかになり、本手法の話題変化に対する有効性が示された。

今後は、全体的な精度向上のため、現段階では当てはめさえも行なわれていない部分への対処を考えていきたい。

参考文献

- [1] 神野, 荒木, 宮永, 枅内: 接続情報を利用した誤りを含む文字列からの原文復元法の性能評価, 信学技報, NLC94-42, pp.25-32 (1994).
- [2] 荒木, 宮永, 枅内: 多段階分割復元法による誤りの多い文字列からの原文の復元, 情処学論, Vol30, No2, pp.169-178 (1989).