7E-3 音声認識候補の正規化認識確率に関する考察

坂野俊哉　北研二　森元逞
ATR自動翻訳電話研究所

1. はじめに

ATRでは自動翻訳電話開発に向けての第一歩として、音声認識処理および言語翻訳処理を統合化した評価システム「SL-TRANS」を開発した[1]。このシステムの音声認識部は、ATRで考案したHMM-LR法により行われている[2]。HMM-LR法は、拡張LRパーザ[3]の中から直接HMM音韻照合モジュールを駆動することにより、音韻ラティス等の中間的な構造を介することなく音声データを解析する手法である。この方法では、図1の様な文節毎に複数の認識候補を出力し、言語翻訳部に渡すため、文節候補群を言語処理可能な範囲まで絞り込む必要がある。筆者らはこれらの候補を絞り込む手法として文節内の候補群の分散および候補間の認識スコアの変化率を利用する方法を提案した[4]。本報告では、認識部で得られた認識候補に付加された認識スコアを言語処理できる正規化確率へ変換する手法と、認識候補群内の正解候補の有無を判定する手法を述べる。

(番号は認識ランク、各候補のカッコ内は認識スコア)

	会議に	参加の	登録
1	会に (1.751883)	参加の(1.529894)	登録 (1.533804)
2	会議 (1.848895)	3かの (1.529894)	登録を(1.683842)
3	会議に(1.878665)	参加を(1.535608)	登録も(1.747440)
4	不利に(1.990288)	3かのを(1.537028)	届く (1.750890)
5	日に (1.995784)	参加のを(1.537028)	登録へ(1.803982)

図1 文節ラティス

2. 統計による候補の絞り込み

図1の認識スコアは、HMMでの各音韻のモデルによって得られた確率値(実際には、その対数値である)の積(対数の場合は和)で与えられており、文節の各候補の認識スコアはある傾向で分散する。次にこの分散による絞り込みおよび2階差分による絞り込みを説明する。

2.1. 分散による絞り込み

各文節候補に付属している認識スコアの分散を階級値とする文節認識率の傾向を、415個の文節からなる音声認識結果から求めた結果、図2に示す分布となった。このうち、認識ランクと分散のグラフを図3に示す。これから各文節毎にその候補群の分散値及び、その正しい候補を包含する最低ランク(完全認識ランク)との分布モデルを以下のように1次近似式として設定できる。

(vは対象文節候補群の分散値、dは分布パラメータを示す。)

$$\mathrm{rank}(v) = -\frac{5}{d}v + 5 \quad (0<d<0.1)$$

この分布モデル式を用いて、131個の文節からなるサンプル(1文節当たりの候補数は5個)を評価した結果は、文節候補削減率53.3%、エラー率4.6%であり、1文節当たりの平均候補数は2.3個にまで減らすことができた。

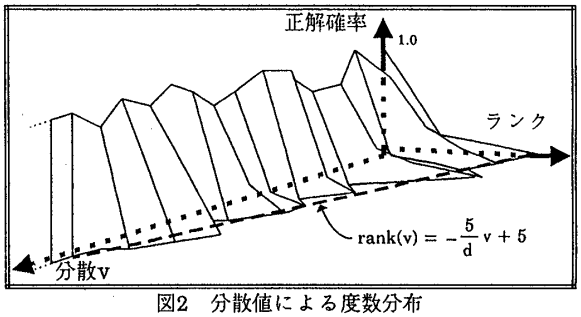

図2 分散値による度数分布

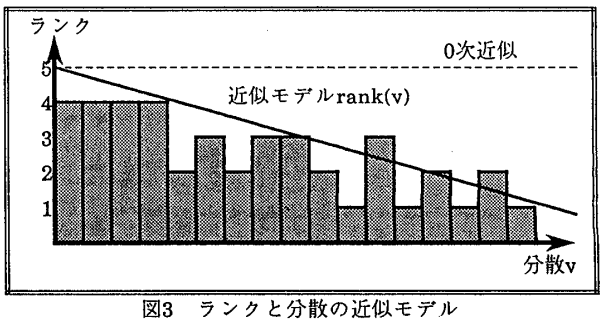

図3 ランクと分散の近似モデル

2.2. 2階差分による絞り込み

データ列$\{D_i\}$($i \geqq 1$)に対して、

$D''_1 = D_2 - D_1$

$D''_i = D_{i-1} - 2D_i + D_{i+1}$ ($i \geqq 2$)

をD_iの2階差分という。

各文節候補群$\{D_i\}$に対して2階差分を求め、その最大値がD''_j ($j \geqq 2$)の時、jをこの文節候補群に対する完全認識ランクとする方法を検討した。文節候補群の最大2階差分値を特性値とし、各々の階級毎にこの方法の成功文節数、失敗文節数の度数を音声認識結果(415文節)から求めたが、最大2階差分値が或る閾値以上に対して成功率は約98%となっている。そこで、最大差分値が閾値に満たないラティスに対しては従来の分散による方法で、それ以外では2階差分による方法で完全認識モデルランクを求めることにする。この方式を先のサンプルデータに適用した結果、文節候補削減率は57.3%、エラー率は1.5%、1文節当たりの平均候補数は2.1個となり、分散による方法単独で適用した結果よりも改善されている。

以下で、分散モデルを利用した認識スコアの正規化と、2階差分値による正解候補存在判定法を述べる。

---

a Study of Normalized Plobabilities for Speech Recognition Candidates
Toshiya SAKANO, Kenji KITA, Tsuyoshi MORIMOTO
ATR Interpreting Telephony Research Laboratories

3. 認識スコアの正規化

前述の通り認識スコアはHMMによる音韻照合の際に、音韻毎に異なるモデルを用いて算出された値であり、そのため文節間では単純に比較できないので、音声認識候補の正解確率値としては使用し難く、なんらかの方法で正規化する必要がある。

図2における分散階級のランクと正解確率のグラフを図4に示す。ランクに対する正解確率は分散階級毎に一定の分布に従っていると思われる。この分布p(v,r)(vは分散階級値、rはランクを表し、p(v,r)自身は連続分布密度を表しており、そのr=0からr=∞までの積分値は1である)を各分散階級の文節候補群に対する正解確率の分布モデルとみなすことができる。しかも、この確率値は任意の分散階級間で正規化されたものとみなすことができるから、各文節候補群の間で任意の候補に対する正解確率の単純比較が可能になり、文節候補列で構成される文候補に対して文としての正解確率が定義できる。例えば、文節群L_1、L_2、L_3に対する各文節候補群の確率分布が各々{x_{11},x_{12},x_{13}}、{x_{21}}、{x_{31},x_{32}}とすると、文候補の確率分布は次のようになる。(この場合、$x_{21}=1$)

$$x_{11} \times x_{21} \times x_{31}$$
$$x_{11} \times x_{21} \times x_{32}$$
$$x_{12} \times x_{21} \times x_{31}$$
$$x_{12} \times x_{21} \times x_{32}$$
$$x_{13} \times x_{21} \times x_{31}$$
$$x_{13} \times x_{21} \times x_{32}$$

この値の大きい文候補から文法や意味などによりチェックする。この方式の利点は、文節候補群内の確率分布の形に依らず、文候補の確からしさが求まることである。

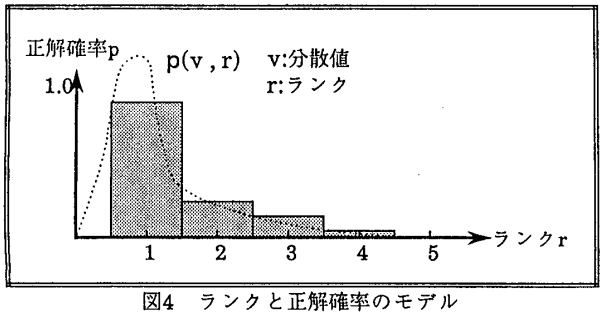

図4 ランクと正解確率のモデル

実際の音声言語統合化システム「SL-TRANS」では、この分布モデルp(v,r)を次のように近似している。即ち、各ランク毎に分散階級値と正解確率の分布の近似式を求め、これをp(v,r)の近似分布として用いている。(図5参照)

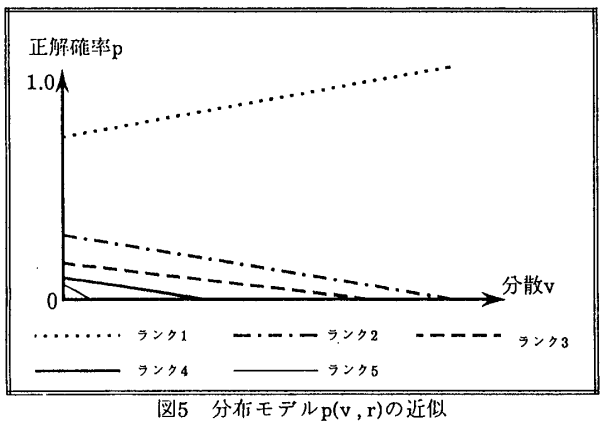

図5 分布モデルp(v,r)の近似

分散階級の正規化正解確率を求めるには、ランク1から5までの近似式からその分散階級に相当する値を標準化することにより求められる。その際、近似式の値が0のランクを削除することにより候補の絞り込みを行う。

「SL-TRANS」の37文(83文節)からなる例文に対してこの候補の絞り込み法を適法すると1文節失敗するため、「SL-TRANS」内では正解確率の正規化のみを行い、次処理で候補の絞り込みを行うことにしている。

4. 正解候補の存在判定法

HMM-LR法の音声認識率は1文節あたり第5位までで約96%に達しており、かなり良い値を得ている。しかし、認識失敗つまり第5位(「SL-TRANS」では、第5位までの結果を出力している)までの候補の中に正解が出てこないこともある。このまま正解のない音声認識結果を言語処理に渡すと正しく言語翻訳されないであろう。この場合、言語処理部に行くまでに正解候補の存在を判定する必要があり、その結果、音声認識候補中に正解候補がないと判断されると、発話者に再発声あるいは発声語句をキーインを要求できる。これは、人が他人の言うことを聞き取れなっかた場合に聞き直すことに似ている。

正解候補の存在判定は、前述の正解確率分布モデルp(v,r)を用いてできる。例えば、ランク第6位までのp(v,r)の値が正値ならば、第6位の音声認識候補が正解である可能性があり、従って第5位までしか出力されない認識候補中に正解がないかもしれず、発話者に問い合わせる必要があると判断される。

また、別の判定法として、2階差分による方法で求めた最大2階差分値を用いる方法があり、次のようにして行っている。完全認識ランクを求めた場合には、2階差分値に対して信頼度の高い方向に閾値を設定したが、今度は信頼度の低い方向に設定する。認識候補群の2階差分値がこの閾値よりも低ければ、認識候補中には正解候補がない可能性があるとみて発話者に問い合わせている。「SL-TRANS」では、この方法を採用しており、その結果、認識されない3個の文節のうち1個に対して正解がないと判定している。

5. むすび

以上、HMM-LR法による音声認識結果に対して文節候補の正解確率を正規化する方法および、正解候補の存在判定法を実際に「SL-TRANS」に適用した例とを併せて説明した。この中で述べた正解確率分布モデルp(v,r)は、HMMあるいはHMM-LR法の構造を反映したものとみなすことができ、今後はこのモデルを精密化することにより、音声認識自身にも役立てていく予定である。

謝辞

研究の機会を与えて頂いたATR自動翻訳電話研究所 樽松明 社長に深謝致します。また熱心に討論して頂いた同研究所データ処理研究室の皆様に感謝致します。

[参考文献]

[1] 森元・小倉・相沢・樽松:「音声言語日英翻訳実験システム(SL-TRANS)の概要」情報処理学会 1989年秋期全国大会

[2] 北・川端・斎藤:「HMM音韻認識と予測LRパーザを用いた文節認識」電子情報通信学会 SP88-88

[3] M.Tomita:"Efficient Parsing for Natural Language - A Fast Algorithm for Practical Systems", Kluwer Academic Publishers(1986)

[4] 坂野・北・森元:「音声認識候補の統計処理による絞り込み」電子情報通信学会 1989年春期全国大会