

木構造を用いた音韻連鎖統計モデル

A Tree-based Stochastic Phone Sequence Modeling

田本真詞

伊藤克亘

田中穂積

TAMOTO masafumi ITOU katunobu TANAKA hozumi

東京工業大学工学部

Tokyo Institute of Technology

あらまし 連続音声認識において音韻認識誤りを補正するための言語情報として、従来から N-gram による音韻連鎖統計モデルが広く用いられている。しかし N-gram モデルでは、音韻連鎖長 N の値が増大するにつれ、モデルの推定精度が向上する反面、モデルの推定に用いるデータ量が一定である場合に逆に信頼性が損なわれることがあった。本研究では、統計モデルの信頼性を損なうことなく推定精度を向上させる手法として、情報量の大きな音韻連鎖を選択的に成長させる、木構造を用いた音韻連鎖統計モデルを提案する。また、実際に N-gram との比較実験を行ない、本モデルの有効性を検証した。

Abstract In continuous speech recognition, N-gram model of phone sequence are widely used for correcting recognition error. By increasing phone sequence length N of N-gram, accuracy of the model improved, on the other hand, reliability of the model decreased with fixed number of training data. In this paper, we propose the tree-based stochastic phone sequence model by selecting the best scored sequence will be expand that improve the accuracy of the statistical model without losing its reliability.

1 はじめに

連続音声認識処理における音韻認識率は次第に向上しているが、認識誤りを生じないような音韻モデルは存在していない。そこで、音響的情報のみによる認識結果を補正し、認識の性能を向上させるための手段として、文脈、意味、文法など様々な言語情報を利用することが考えられる。これらの情報のひとつとして、音韻連鎖に関する統計情報が注目されている [中川 88]。実際、音韻連鎖の統計的言語モデルが音声認識に有効に働くことが知られている [川端 90]。このような統計モデルの抱える問題として、精度と信頼性の間にトレードオフがあることが知られている。すなわち、モデルの扱う音韻連鎖を長くしてモデルの組合せを増大させ、分解能の高い、精度の高いモデルを実現すること、各音韻連鎖の組合せについて、統計的情報を抽出するために十分な標本数を確保し、モデルの信頼性を損なわないことは、両立が難しい。

本研究では、このトレードオフを軽減させるための手法として、木構造を用いた統計モデルを提案する。このモデルでは、音韻のコンテキストに応じて参照する連鎖の長さを動的に変化させ、信頼性を損なうことなく音韻連鎖モデルの精度を向上させることを目指す。

コンテキストの長さを固定せず、コンテキストの持つ情報量に注目して連鎖を分割、成長させる手法が、英単語の

出現確率を推定する 2 分木モデルの生成に有効であることが報告されている [Bahl 89]。

また、日本語の音韻連鎖において、長母音を考慮することで実質的に音韻連鎖を伸長した場合に、音韻連鎖の認識率が向上することから、モデルを推定するデータ中の連鎖パターンの頻度をもとに、出現確率の高い音韻連鎖を選択的に成長させる方法がモデルの精度向上に貢献するとの提案がなされている [伊藤 90]。

これらの研究をふまえ、モデル推定に用いるデータのタスクに対応して木構造を自動的に成長させるアルゴリズムを検討し、実際の推定データをもとに木構造モデルを生成する。さらに、この木構造モデルと N-gram を比較する実験を行ない、本モデルの有効性を検証する。

2 音韻連鎖の統計モデル

ある言語において文を表す音韻連鎖 $S = p_1, p_2, \dots, p_n$ が生成される確率 $P(S)$ は、次の式で表すことができる [中川 88]。

$$P(S) = \prod_{i=1}^n P(p_i | p_1, p_2, \dots, p_{i-1}) \quad (1)$$

ここで $P(p_i | p_1, p_2, \dots, p_{i-1})$ は、音韻連鎖 p_1, p_2, \dots, p_{i-1} が生成された後に、音韻 p_i が生成される条件付き確率であ

る。ところが、一般に文の長さには制限がないため $P(p_i | p_1, p_2, \dots, p_{i-1})$ を求めることはできない。仮に文が有限長の音韻連鎖で構成されるとしても、 p_1, p_2, \dots, p_{i-1} の組合せは、音韻連鎖に関して指数的に増大する。よって個々の音韻連鎖について確率値を与えることは現実的ではない。そのため、音韻連鎖の統計モデルを推定するには、何らかの手法で $P(p_i | p_1, p_2, \dots, p_{i-1})$ を近似する必要がある。これらの手法のひとつに N 個の音韻連鎖の出現確率 (N -gram) を用いた近似がある。

2.1 N-gram モデル

音韻連鎖 $S = p_1, p_2, \dots, p_n$ が生成される確率は、 N -gram を用いて近似的に次のように表せる。

$$P(S) = \prod_{i=1}^n P(p_i | p_{i-N+1}, p_{i-N+2}, \dots, p_{i-1}) \quad (2)$$

この近似方法では、 N の値が増大するにつれ、考慮すべきコンテキストの種類が増え、推定精度が向上すると考えられる。ところが N を大きくすると先に述べたように $p_{i-N+1}, p_{i-N+2}, \dots, p_{i-1}$ の組合せが指数的に増加する。 N -gram の表す生成確率は、データ中の出現頻度から算出するので、ある一定量のデータから生成確率を算出する場合、各コンテキストについて統計的情報を十分に抽出できるだけのサンプル数を確保できないと逆に統計モデルとしての信頼性が損なわれ、かえって認識率を低下させることになる。

また、 N が大きくなると統計モデルにタスクによる相違が生じ、robustness が失われることが知られている [伊藤 90]

このような問題に対して、削除補間法による N -gram の線形結合 [Jelinek 80] や、出現頻度の小さい N -gram 確率の back-off スムージング法による推定 [Katz 89] 等のデータ補間を行なって精度と信頼性を向上させる試みがなされている。しかし、近年、統計的言語モデルの分野で注目されている $N = 3$ の連鎖確率 (トライグラム) に代わる決定的な方法は見い出されていない [Jelinek 91]。

2.2 木構造モデル

これまで述べたようにコンテキスト依存性の高い、すなわち音韻連鎖の長い統計モデルでは、特定のコンテキストにおける音韻の生成確率をより精密にモデル化することができる。したがってモデルの推定精度は高いが、信頼性を維持するためには大量の推定データが必要になる。逆にコンテキスト依存性の低い、音韻連鎖の短いモデルでは、生成確率が様々なコンテキストの混合となるため精密さに欠けるが、少量のデータで信頼性の高いモデル化ができる。

そこで、音韻連鎖の生成確率を効率良くモデル化するためには、音韻連鎖のコンテキストの持つ統計的な性質を考慮する必要がある。直観的には出現頻度の大きいコンテキストの音韻連鎖長 N をより大きくすれば、全体の信頼性を損なうことなく、精度の高いモデル化が期待できる。

ここでコンテキストによって出現頻度に差が生じる理由について考えてみると、日本語の音韻的な特性から、文中に頻出する付属語や漢字など語構成要素 [水谷 87] によって、長さ 3 ないし 4 の音韻連鎖がある程度決定されている

ことがあげられる。このことから逆に、語構成要素のレベルで決定される音韻連鎖は、局所的なコンテキストに強く依存することが推定できる。つまり、ある音韻の生成確率を求めるには、その直近、すなわち語構成要素の範囲の音韻連鎖がわかればよいことになる。

以上のことから、局所的な音韻連鎖のコンテキストに応じて連鎖長を自由に変化させ、局所的な音韻連鎖の依存性を考慮したデータ構造として、木構造を用いたモデル化を行なうこととした。

3 木構造を用いた音韻連鎖の統計モデル化

3.1 木構造による音韻連鎖の表現

コンテキストによって連鎖長が動的に変化する N -gram を表現するには、次のような構造を想定すればよい。まず、

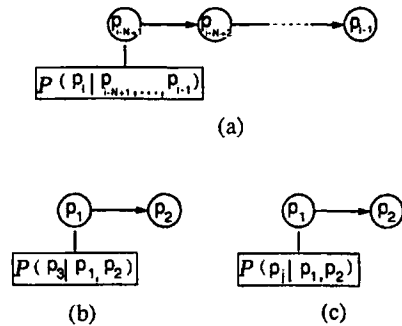


図 1: 木構造による音韻連鎖の表現 1

$P(p_i | p_1, p_2, \dots, p_{i-1})$ を音韻を表すノードと音韻同士の結合を表すエッジで構成されるグラフで表す図 1(a)。具体的に 3-gram (trigram) を例にとって考える。 $P(p_3 | p_1, p_2)$ を、音韻列 p_1, p_2 (演算子 \cdot は音韻 (列) の結合を表す) の後に音韻 p_3 が出現する条件付き確率とする ($\{p_i | i = 1, 2, \dots, N\}$ は、音韻連鎖の順序ではなく、音韻の集合を表す) 図 1(b)。図中の $P(p_3 | p_1, p_2)$ は、ノードに確率値が割り当てられていることを表す。

これは、 p_3 に任意の音韻を割り当てることで図 1(c) のように一般化できる。

次に音韻の一般化をノードに割り当てられた音韻に対して適用する。 p_1 の部分に任意の音韻が出現した場合の音韻連鎖の生成確率は、グラフを分岐させることで次のように表せる。同じように全てのノードについて音韻ごとにグラフを分岐させることで右のような木構造が表現できる図 2。

さらに音韻連鎖長 N を任意とすると、ある音韻連鎖の任意の部分における音韻の生成確率は、音韻連鎖の先端だけでなく、連鎖の中間を用いて表現することができる図 3。

さらに、図 2 と図 3 を組み合わせることで任意の長さのコンテキストにおける音韻の生成確率を表す木構造を構築することができる図 4。

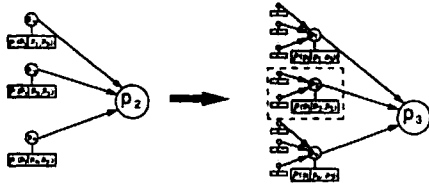


図 2: 木構造による音韻連鎖の表現 2

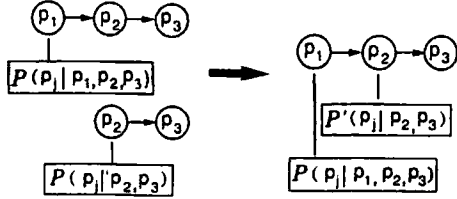


図 3: 木構造による音韻連鎖の表現 3

3.2 木構造の生成手続き

次に木構造を自動的に成長させる手続きについて検討する。伸長させる音韻を決定する戦略には、音韻連鎖の出現確率を連鎖長で正規化して最大の葉(リーフ)ノードを選択する方法、分割前後の相互情報量を最小化する方法などが考えられるが、統計モデルの信頼性の高さを保ちながら精度を向上させ、さらに音韻認識時の確率的言語情報として木構造を利用することを考慮して、木構造の持つ情報量の増加が常に最大になるようにリーフの追加を行なう。これは、木に含まれるすべてのノードについて新たなノードを追加した後の情報量の増分を求め、これを最大にするノードを探索することに等しい。

音韻連鎖を伸長させる手続き

木構造の初期状態は、先行する音韻のないことをしめす①が記入されたルートノードだけが存在する図 5(a)。伸長される音韻を決定する手続きを任意のコンテキストに対する手続きに一般化して、あるリーフに新たなリーフを追加する手続きを説明すると次のようになる。

1. ノードの表すコンテキストが l_i に相当するようなリーフ l_i を考える。ルートノードならば l_i の長さは 0 であ

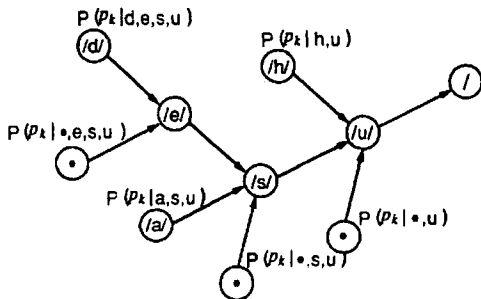


図 4: 音韻連鎖の生成確率 4

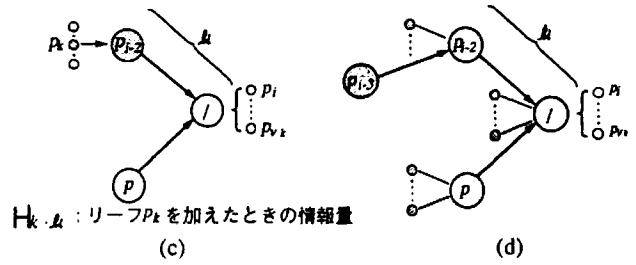
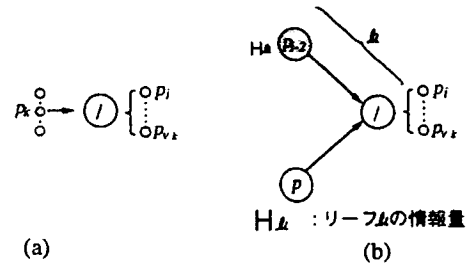


図 5: 音韻連鎖を伸長させる手続き

る。このノードの固有な情報量を H_{l_i} とする図 5(b)。

$$H_{l_i} = \sum_j -P(l_i \cdot p_j) \cdot \log P(l_i \cdot p_j) \quad (3)$$

2. リーフを追加したときの情報量の増加分を計算する図 5(c)。

(a) 新たに追加されたリーフに相当する音韻を p_k としたとき、 p_k の伸長によって得られる情報量 $H_{p_k \cdot l_i}$

$$H_{p_k \cdot l_i} = \sum_j -P(p_k \cdot l_i \cdot p_j) \cdot \log P(p_k \cdot l_i \cdot p_j) \quad (4)$$

(b) リーフの追加により、 l_i の持つ情報量が変化する。すなわち l_i のもつ情報量は、音韻連鎖 l_i に伸長された音韻 p_k 以外の音韻が先行する時の情報量に場合わけされる。変化後の情報量を H'_{l_i} とする。

$$H'_{l_i} = \sum_j -P(p_{x \neq k} \cdot l_i \cdot p_j) \cdot \log P(p_{x \neq k} \cdot l_i \cdot p_j) \quad (5)$$

リーフ l_i に音韻 p_k を先行させた場合の情報量の増分は $H'_{l_i} + H_{p_k \cdot l_i} - H_{l_i}$ で求められる。

3. リーフ l_i において各音韻 p_k ($k = 1 \dots n$) を伸長した時の情報量の差分を求める。同様にすべてのリーフノード l_i ($i = 1 \dots N$) における各音韻ごとの情報量の差分を求める。
4. すべてのノードについて、すべての分割に関する情報量の増分を比較し、増分が最大となるノードを新たに追加する図 5(d)。
5. 新しいリーフノードを l_{i+1} とし、1. にもどる。

3.3 音韻連鎖の生成確率

生成された木構造モデルにおける音韻連鎖の生成確率は、次のように表される。

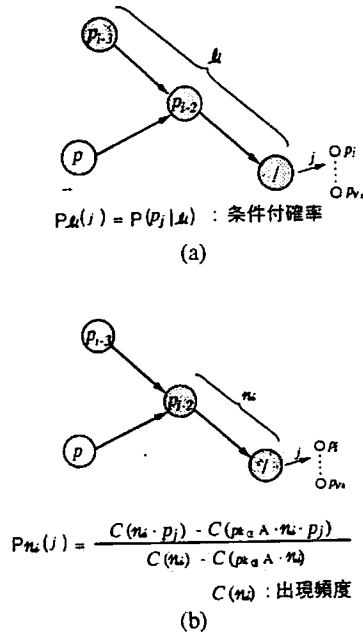


図 6: 音韻連鎖の生成確率

リーフノードにおける生成確率

リーフノードの表すコンテキストを l_i とするとき、 l_i に続いて音韻 p_j が生成される確率 $P_{l_i}(j)$ は、N-gram で表される条件つき確率に等しい図 6(a)。

$$P_{l_i}(j) = P(p_j | l_i) \quad (6)$$

リーフ以外のノードにおける生成確率

リーフ以外のノードにおける生成確率は、音韻が後方に伸長している分だけ変化する。この値は条件つき確率だけでは求められない。ノード n_i において先行する音韻の集合が $p_{k \in A}$ であるとき、音韻連鎖 $p_{k \in A} \cdot n_i \cdot p_j$ の生成確率は、学習データにおける音韻連鎖 P の出現回数を $C(P)$ とすることで次のように表せる図 6(b)。

$$P_{n_i}(j) = \frac{C(n_i \cdot p_j) - C(p_{k \in A} \cdot n_i \cdot p_j)}{C(n_i) - C(p_{k \in A} \cdot n_i)} \quad (7)$$

4 木構造の生成実験

大規模なテキストデータベースを利用して任意の音韻連鎖の出現頻度を計数し、このデータをもとに木構造モデルを生成する実験を行なった。

4.1 実験データ

実験に用いたテキストデータベースは、1982年の日本経済新聞の記事 37 日分の 15,207 文、文節数は 145,718、音素

数で 1,385,082 である。これらのデータは、文節区切り、文境界が示された形になっている。そこで、テキストデータベースから音韻連鎖の頻度情報を抽出する時は、音韻連鎖が文節境界の無音区間を含みながら文節間を渡ることを許可したが、文境界を越えて頻度情報を抽出することはない。

また、生成された言語モデルの評価用として、同新聞の 1 日分の記事、571 文、6260 文節、54733 音韻を推定用のデータとは別に用意した。

4.2 モデルの評価方法

生成された統計モデルの評価尺度として以下に述べるような方法を用いる。

coverage

統計モデルによって、テストセットの生成確率を算出する場合、参照されたモデルのうち、生成確率が 0 とならなかったモデルが全体に占める割合を % で表す。この値が大きいほどモデルが訓練データのタスクを良く表現できているといえる。

音韻の平均生成確率

音韻連鎖が $P = p_1, p_2, \dots, p_{N-1}$ である一文に含まれる音韻の平均生成確率を次の式で定義する。

$$P(P) = \prod_{i=1}^n P(p_i | p_{i-N+1}, p_{i-N+3}, \dots, p_{i-1}) \quad (8)$$

これは、一文中のある音韻の生成 (出現が予測される) 確率の平均値である。したがって、精度・信頼性の高い言語統計モデルほど「日本語らしさ」を持った文に対して高い値を示す。そこで文集合に対する生成確率の平均値を言語統計モデルの精度・信頼性の高さを示す尺度のひとつとして評価する。

パープレキシティー

パープレキシティーは、統計モデルを連続音声認識処理で用いる確率的言語情報としたときの、認識のタスクの複雑さを評価する尺度のひとつである。これは、認識対象のエントロピーをもとに算出される、情報理論的な意味での平均分岐数である。例えば、言語 L の音韻を単位としたときのエントロピーが $H(L)$ ならば、次の音韻を決定するのに $2^{H(L)}$ 回の選択が必要になる。言い換えれば $2^{H(L)}$ 個の音韻から 1 音韻を選び出すことになる。

パープレキシティーは、値が大きくなるほど複雑さが増大し、認識が困難になる。

認識対象となる言語 $L = \{w_k | w_k = w_{i1}w_{i2} \dots w_{ik}\}$ のエントロピー $H_0(L)$ は次式のように求められる。

$$H_0(L) = - \sum_{w_k} P(w_k) \log_2 P(w_k) \quad (9)$$

また、1 音韻あたりのエントロピー $H(L_0)$ とエントロピーをもとに算出したパープレキシティー $F_P(L)$ は次のよう

に求められる [中川 88]。

$$H(L) = - \sum_{w_k^i} \frac{1}{k} P(w_k^i) \log_2 P(w_k^i). \quad (10)$$

$$F_p(L) = 2^{H(L)}. \quad (11)$$

また、パープレキシティの値は、認識評価に用いる文に依存するので、ある程度の大きさを持った文集合に対するパープレキシティを求める必要がある。これをテストセットパープレキシティとする [村瀬 90]。

5 結果

5.1 生成された木構造モデル

4.1 節で述べたテキストデータをモデル推定データをもとに 3.2 節で説明したアルゴリズムを用いて木構造モデルの生成を行なった。モデルの成長は、音韻連鎖の伸長・分割によって生じるノードが 10,000 になるまで繰り返した。生成された木構造モデルの一部を図 12 に示す。この時生成されたモデルの数は、430,000 そのうち生成確率が 0 の (flooring される) モデルは、326,811 (76.0%) であった。

生成されたモデルの音韻連鎖の特徴として 2.2 で述べた語構成要素が数多く出現することがあげられる。この理由について考えると、分割・伸長の条件として、木構造全体の情報量を最大化する戦略をとったが、これは結果的に個々の音韻連鎖の持つ情報量を最大化したことになる。つまり、ある音韻列を伸長させたときに情報量がかえって減少してしまう場合は、音韻列の成長は、その時点で終了する。ここで、このような音韻列の持つ特徴を考えると

1. この音韻連鎖に後続する音韻の分布は、一様である。
2. 音韻連鎖に含まれる個々の音韻は、結合性が高い、なぜなら音韻の伸長によって後続する音韻の分布が変化するなら (分布の一様性が損なわれるなら) 情報量が減少し、連鎖の伸長がおさえられてしまう。
3. 後続する音韻の分布が一様であるためには、音韻連鎖自体、ある程度のサンプル数 (出現数) が確保できなければならない、よって頻出する音韻連鎖である。

出現頻度、結合性の高さや、後続音韻の一様性は、頻出名詞や活用語の語幹などの語構成要素に一致する。

この木構造は、音韻/u/で終る連鎖のうち、生成された順に 100 パターンを抽出したものである。図中/#/は、文節間の区切りを表す記号 (ポーズ) である。生成確率は、図左側のリーフノードだけではなく、図 4 で示される通り、中間ノードにも記憶される。図では、/e/--/k/--/o/--/k/--/u/, /k/--/a/--/k/--/u/, /s/--/u/--/r/--/u/などの瀬出語、語構成要素が見られる。この結果から、木構造全体のエントロピーを最大化する方法によって、モデル推定データのタスクを反映した統計モデルが自動的に生成できることがわかった。

```

/#/--/b/--/u/
/#/--/k/--/a/--/k/--
/#/--/k/--/o/--
|#/--/m/--
/#/--/s/--/u/--/r/--
|#/--/sh/--
/#/--/y/--
/N/--/s/--
/N/--/s/--/u/--/r/--
/a/--/#/--/a/--
/a/--/#/--/h/--
/a/--/#/--/s/--
/a/--/#/--
/a/--/b/--
/a/--/e/--/r/--
/a/--/i/--/s/--
/a/--
/a/--
/a/--/z/--
/e/--/k/--/o/--/k/--
/e/--/a/--/r/--
/e/--/i/--
/g/--/a/--/k/--
/g/--
/h/--/a/--/ts/--
/h/--/u/--/k/--
/i/--/#/--/ts/--
/i/--/b/--
/i/--/h/--
/i/--/j/--
/i/--/k/--/a/--/k/--
/i/--
/i/--/m/--
/i/--/s/--/a/--/k/--
/i/--/s/--/u/--/r/--
/i/--/z/--
/j/--/i/--/ts/--
/k/--/a/--
/k/--/e/--/r/--
/k/--/i/--
/ky/--/o/--/k/--
/m/--/e/--/r/--
/n/--/a/--/k/--
/n/--/a/--/r/--
/o/--/s/--
/o/--/s/--/u/--/r/--
/o/--/#/--/h/--
/o/--/#/--/k/--
/o/--/#/--/s/--
/o/--/#/--/t/--
/o/--/#/--
/o/--
/r/--/e/--/r/--
/r/--/i/--/ts/--
/ry/--/o/--/k/--
/s/--/e/--/ts/--
/s/--/o/--/k/--
/sh/--
/sh/--/u/--/ts/--
/t/--/a/--/k/--
/t/--/o/--
/u/--/s/--
/u/--/sh/--
/y/--/a/--/k/--

```

図 12: 生成された木構造モデルの例

5.2 評価

評価実験の対象として木構造モデルを生成したものと同一テキストデータを使い、 $N = 2, \dots, 5$ の N-gram モデルを生成した。この N-gram モデルと木構造モデルに対して 4.1 節で示した評価データを与え、4.2 節の各評価を行なった。

評価は、与えられた評価文に対して適用できるモデルを検索することから始まる。

5.3 生成確率参照のアルゴリズム

生成された木構造モデルを用いて、ある音韻の生成確率を参照することは、音韻連鎖 p_1, p_2, \dots, p_{n-1} の後に任意の音韻 p_n の続く条件つき確率を参照する点で N-gram と同一であるが、音韻連鎖の長さが文脈によって変化する点が大きく異なる。

モデルの構造を利用した探索法として、音韻連鎖を前方から後方へ走査しながら、木構造をルートノードからリーフへ探索するアルゴリズムが考えられる。

生成確率を参照する手続き

1. 木構造上の位置を表すポインタをルートノードに、検索する音韻を示すカーソルを音韻連鎖 p_1, p_2, \dots, p_{n-1} の最前方 (p_{n-1}) に合わせる。
2. カーソルで示される音韻がポインタで示されるノードからリーフ方向に伸長しているかを調べる。

音韻が伸長している時

ポインタを伸長しているノードに移動し、カーソルを一音韻分、後方へ進める。

音韻が伸長していない時

- (a) ポインタがリーフの位置にある時
ポインタの示しているノードの確率を結果として参照する。
- (b) ポインタがリーフ以外のノードの位置にある時
木構造はカーソルの示している音韻以前の音韻連鎖に関する生成確率しか持っていない。この場合、カーソルの示している音韻は、木構造の生成時、音韻連鎖の伸長の対象にならなかった特別な音韻 ('*') として扱われる。'*' を連鎖の最後尾に持つ音韻連鎖の生成確率は、(7) 式を用いることで、リーフ以外の全てのノードで参照できる。

例えば SilB/k/o/r/e/d/e/s/u/SilE (SilB は、文頭を、SilE は、文末を表す) という文を 3-gram と SilB/k/, /k/o/r/e/, /d/e/s/u/ という音韻連鎖を持つ木構造モデル (Vari-gram) で音韻列を検索した時の違いを見ると次のようになる。3-gram(N-gram) では、コンテキストによる音韻連鎖の違いはなく常に一定であるが、Vari-gram では、コンテキストに応じて、参照する音韻連鎖の長さが変化する。

SilB	
/k/	$P(/k/ \text{SilB})$
/o/	$P(/o/ \text{SilB}/k/)$
/r/	$P(/r/ /k/o/)$
/e/	$P(/e/ /o/r/)$
/d/	$P(/d/ /r/e/)$
/e/	$P(/e/ /e/d/)$
/s/	$P(/s/ /d/e/)$
/u/	$P(/u/ /e/s/)$
SilE	$P(/SilE/ /s/u/)$

3-gram

SilB	
/k/	$P(/k/ \text{SilB})$
/o/	$P(/o/ /k/)$
/r/	$P(/r/ /k/o/)$
/e/	$P(/e/ /k/o/r/)$
/d/	$P(/d/ /k/o/r/e/)$
/e/	$P(/e/ /d/)$
/s/	$P(/s/ /d/e/)$
/u/	$P(/u/ /d/e/s/)$
SilE	$P(/SilE/ /d/e/s/u/)$

Vari-gram

図 7: 音韻連鎖の参照

る。これにより瀕出する音韻連鎖は、連鎖長を長くして精度を高め、出現率の低い音韻は、短い音韻連鎖で評価し、信頼性を損なわずに推定ができる。

このとき、評価で参照したモデルの平均連鎖長をグラフ 8 に示す。グラフでは、N-gram の各モデルを点でプロットし (グラフの左から 2-gram, 3-gram, 4-gram, 5-gram)、Vari-gram をグラフの曲線で表している。

グラフでは、同じモデル数の Vari-gram と N-gram では、平均連鎖長がほぼ同じか Vari-gram が若干優れていることが示される。

coverage

各モデルの coverage をグラフ 9 にしめす。N-gram では、一般にモデル数の多い (連鎖長 N が増える) ほど同じモデル推定データでもデータ量が相対的に不十分になる。特に出現頻度の低いコンテキストをモデル化した際に coverage が急激に低下する。Vari-gram では、頻度の多いコンテキストからモデル化するので急激な coverage の低下は、見られない。

音韻の平均生成確率

平均生成確率は、図 7 で参照した確率の積を音韻の連鎖長で正規化したものであるが、これもモデルの連鎖長が長くなるにつれ、正確な生成確率が推定でき、確率は上昇する。しかし、N-gram では、coverage の低下などにより、モ

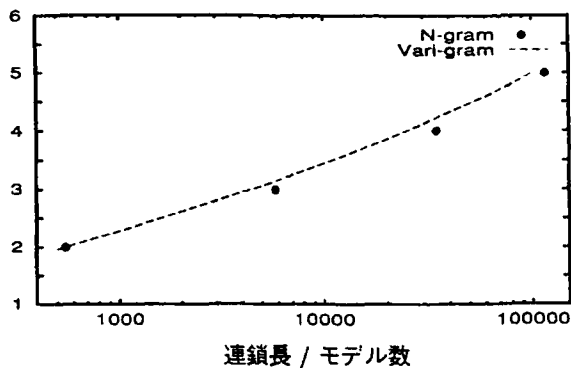


図 8: モデルの平均連鎖長

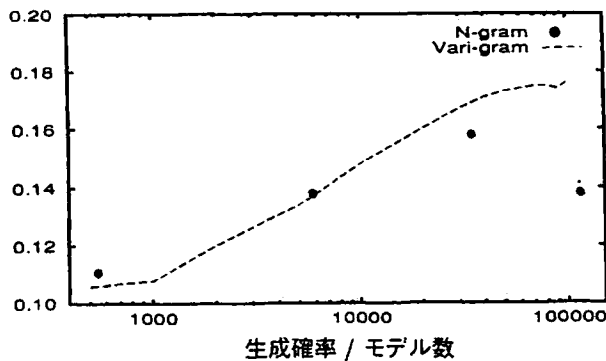


図 10: 音韻連鎖の生成確率

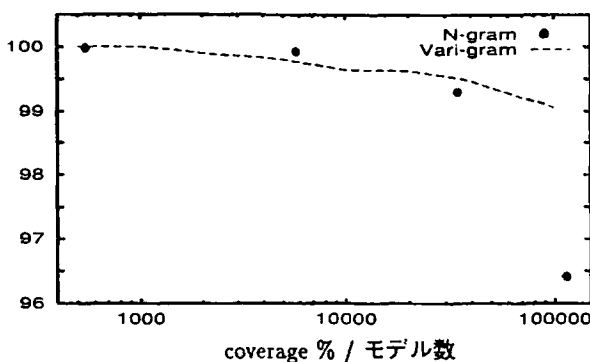


図 9: coverage

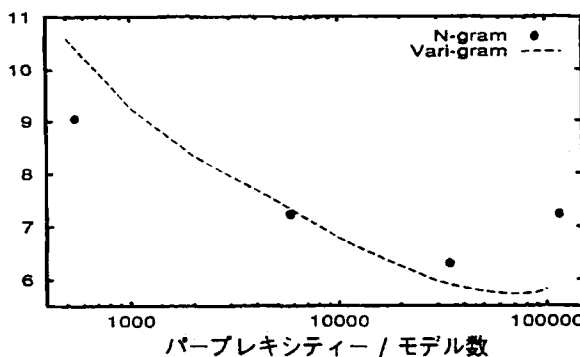


図 11: パープレキシティー

デルの精度が損なわれ生成確率がかえって低下している。Vari-gram では、増加傾向が続いていることがわかる。

テストセットパープレキシティー

認識のタスクの複雑さを表すパープレキシティーも、平均生成確率と同じようにモデルの信頼性の低下とともに値が増加する。Vari-gram も 100,000 モデル付近でパープレキシティーが増加しているが、このモデル数は木構造生成に用いたデータ量に依存する限界値と推察できる。

6 まとめ

木構造の持つ情報量が最大になるよう、音韻連鎖のコンテキストに応じて参照する音韻連鎖の長さを動的に変化させ、モデルの持つ信頼性を損なわずに精度を向上させる音韻連鎖の統計モデルを提案した。

この木構造は、先行する音韻をさらに伸ばした時に生じる情報量が最大になるような音韻を選択し、少ないモデル数で大きな情報量を蓄積できる構成をとっている。

1,385,082 音節からなるテキストデータベースをもとに生成した木構造モデルは、同じデータから生成した N-gram に対して、テストセットに対する coverage、音素の平均生

成確率、パープレキシティーなどの評価で優れていることが示された。

特に

1. 出現頻度の高いコンテキストを選択的に成長させるため、N-gram にみられるように、連鎖長が大きくなっても coverage が急激に低下することがなく robustness に優れている。
2. 統計モデルの精度と、そのモデルを推定するためのデータの規模に関する定量的な比較実験として、N-gram と Vari-gram の推定データの規模に対するモデル化能力の差の有無を試すことができる。先に述べたように木構造モデルは、タスク適応化能力に優れているため、N-gram より効率的にコンテキストをモデル化できる。
3. 木構造モデルの生成する音韻連鎖を観察すると、「ギジュツ」「キカク」「ケンセツ」「キョーリョク」など、経済系の新聞記事のタスクに出現しがちな単語が多くみられる。このことから、情報量最大化による木構造の生成は、タスク抽出能力が大きいと考えられる。

ここで、木構造モデルを実際の音声認識における言語情報として用いる場合について

1. 平均生成確率が高いことから、認識時の正解候補を枝刈りさせにくくする。

2. パープレキシティーの低さから、認識処理における候補数が削減されタスクが軽減される。

ことが考えられる。また、問題点として

1. 生成されたままの木構造モデルでは、音韻を前方から後方へ(時間軸とは逆方向に)探索するので音声認識などのリアルタイム処理に向かない。

2. テキストデータベースを更新した場合、モデルも新規に再構築しなければならない。木構造モデルでは、音韻を伸長・分割する際にクラスタリングなどの処理を行わないため、計算コストは比較的低い。しかし、モデルの再構築は、テキストデータが大量になるほど処理に多くの時間が必要になるので、インクリメンタルなモデルの再構成を実現することが望まれる。

ことがあげられる。

木構造を用いた音韻連鎖統計モデルは、情報量最大化とは異なる成長条件に対しても適用可能であり、今後の研究によってより良い条件の発見が期待される。

謝辞

本稿で使用した日本経済新聞の記事に関するテキストデータベースは、NTT 情報通信処理研究所メッセージシステム研究部から提供していただきました。貴重なデータを使用させていただいたことを深く感謝いたします。

また、日頃から研究について御助言、御討論いただく田中研究室の皆さんに感謝いたします。

参考文献

[中川 88] 中川聖一、“確率モデルによる音声認識”、電子情報通信学会 (1988)

[村瀬 90] 村瀬功、上田佳央、中川聖一、“文脈自由文法と bigram・trigram による言語のモデル化の検討”、信学技法、SP-90-75 (1990)

[川端 90] 川端、花沢、伊藤、鹿野、“HMM 音韻認識における音節連鎖統計情報の利用”、信学技法、SP-89-110 (1990)

[水谷 87] 水谷 静夫、田嶋 一夫、佐竹 秀雄、野村 雅昭、石井 雅彦、樺島 忠夫、“文字・表記と語構成”、朝倉日本語新講座、(1987)

[Jelinek 80] F.Jelinek, R.Mercer, “Interpolated Estimation of Markov Source Parameters from Sparse Data”, *Pattern Recognition in Practices*, E.S.Gelsema and L.N.Kananl, ed., North-Holland Publishing Company (1980)

[Katz 89] S. Katz, “Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer”, *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*. ASSP-34(3) 1989

[Jelinek 91] FREDERICK JELINEK, “UP FROM TRIGRAMS!” (1991)

[伊藤 90] 伊藤克巨、“日本語の統計的な振舞いを利用した連続音声認識”、修士論文、東京工業大学、(1990)

[速水 90] 速水 悟、田中 和世、“木構造音韻モデルによる未知音素文脈中の音響的変動の予測と評価”、電子通信情報学会、SP90-64 (1990)

[Bahl 89] Lalit Bahl, Peter F. Brown, Peter V. Souza, “A Tree-Based Statistical Language Model for Natural Language Speech Recognition”, *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*, VOL. 37, NO.7, JULY 1989