

外 3-43

早稲田大学大学院理工学研究科

博 士 論 文 概 要

論 文 題 目

A Study on Language Modeling
for Speech Recognition

音声認識のための言語モデルに関する研究

申 請 者

北 研二

Kenji Kita

平成 4 年 2 月

音声認識の最大の目的は、人間にとて最も自然で快適な意思伝達の手段である音声を、人間と計算機との間のコミュニケーションとして用いようというものである。近年、ニューラル・ネットワークや隠れマルコフ・モデル（HMM）に基づいた音響モデルの研究が著しく進展し、これらに基づいた大語彙連続音声認識システムも構築されるようになった。しかし、いかに精密な音響モデルを用いても、言語的な情報なしでは、高精度の音声認識システムを構築するのは不可能であると考えざるをえない。いくつかの実験から、人は言語情報を用いて音声の聞き取りを行なっており、言語情報がないと連続音声中の音韻を必ずしも正しく聞き取れないことが示されている。従って、音声情報と言語情報を統一的な観点から扱う音声認識の研究が重要であると考えられる。本論文では、言語情報のモデル化の手法として、統語的言語モデルと確率的言語モデルの2つを用い、これらのモデルを音声認識においていかに有効に活用するかについて論じる。

音声認識は、一つの発話に対して、その発話内容を表す音韻列あるいは単語列の仮説を見つけるという探索問題としてモデル化することができる。統語的言語モデルは、数多くの仮説のうちから、言語の文法にかなった仮説だけを選び出すことを可能にし、音声認識の探索範囲を縮小するのに有効なモデルである。正規文法、あるいはそれと等価な有限状態オートマトンは、制御しやすいという利点があるため、多くの音声認識システムで用いられているが、正規文法は自然言語の言語現象を記述するには弱いモデルである。このため、本論文では、統語的言語モデルとして文脈自由文法を用いる。文脈自由文法を用いる利点は、まず第一に多くの言語現象が文脈自由文法で記述されるということがあげられる。第二の利点として、文脈自由文法に対しては、効率的な構文解析アルゴリズムが知られていることがある。多くの構文解析アルゴリズムが開発されているが、その中でも拡張LRアルゴリズムは自然言語の文法を処理するのに最も効率的なアルゴリズムとして知られており、本論文ではこのアルゴリズムを用いる。

また、音声認識では、音響的に似通った数多くの仮説の中から最も確からしい仮説を選択しなければならないが、このためには仮説を定量的に評価する必要がある。通常は、各仮説と入力音声のマッチングのよさを表す音響的な尤度が用いられているが、仮説をより精密に評価するためには、言語的な側面からみた尤度も考慮する必要がある。確率的言語モデルは、仮説の言語的な尤度を計算するのに効果的なモデルである。本論文では、日本語の音節連鎖確率を用いたモデルと、文脈自由文法に確率情報を組み込んだモデルを用いる。

本論文では、音響モデルとして隠れマルコフ・モデル（HMM）を一貫して用いている。HMMは発声状況やコンテキスト等の違いによる音声のゆらぎを統計的に表現できるという特徴があり、音声の変動に対して強いモデルを構成することができる。従って、調音結合の影響が大きい連続音声の認識に有効である。また、認識単位を音韻に設定しておけば、これを基に任意の単語モデルを合成することができるため、大語彙音声認識に適している。

本論文は、本文7章と付録2章から構成されており、以下にその概要を述べる。

第1章では、本論文で扱う2つの言語モデル（統語的言語モデルと確率的言語モデル）を概括する。

第2章では、音韻ベースのHMMと拡張LRアルゴリズムを統合化した音声認識—HMM-LRアルゴリズム—to導入する。拡張LRアルゴリズムは、プログラミング言語処理系の分野でよく用いられているLRアルゴリズムを拡張したものであり、一般的な文脈自由文法を取り扱うことができる。

これは表駆動型の構文解析アルゴリズムであり、バーザ（解析系）の動作を記述してある解析表を参照しながら、入力記号列を左から右にパックトラックなしに解析することができる。解析表は、文脈自由文法から自動的に生成され、バーザの動作を決定するのに用いられる。通常のテキストの解析では解析表を用いて入力記号列を解析するが、HMM-LRアルゴリズムでは入力記号列を与えて解析を行なうのではなく、逆に解析表によって次にくる音韻の予測を行ないながら音声認識を行なう。認識動作の概略は以下のようである。まずLRバーザは解析表を参照して文法上で次に接続しうる（文頭であれば文頭にくることが可能な）音韻を予測する。また、LRバーザは各音韻の継続時間長の最小値、最大値等の統計情報を保持しており、これをもとに各音韻ごとに照合すべき音声区間を決定する。照合する音声区間は、現在までに照合してきた音韻列と予測された音韻の最小継続時間長の和を始端、最大継続時間長の和を終端とする区間である。このように照合すべき音韻と照合区間を決定したのち、HMM音韻照合部を駆動して音韻の照合スコアを求める。照合スコアとは、具体的には、照合音声区間が音韻モデルから生成される尤度であり、forwardアルゴリズムあるいはViterbiアルゴリズムによって計算される。照合スコアがある閾値よりも高いすべての音韻に対して、並行して音韻連鎖の枝を伸ばしていく。実際には音韻連鎖の枝を伸ばす過程において解析する候補の数が増加していくので、照合スコアによるビームサーチを行ないながら認識を進める。照合がすべて終った段階で照合スコアの高い第n候補までを認識結果として出力する。このように、HMM-LRアルゴリズムでは音韻ラティスなどの中间形式を介さず認識が進むので、照合スコアの計算が正確になり高い認識性能が得られる。

第3章では、HMM-LRアルゴリズムに基づいた音声認識システム—HMM-LR音声認識システム—のインプリメンテーションとシステムの評価について述べる。まず最初に、音声の特徴量としてスペクトルだけを用いた單一コードブック型HMMを使った、日本語の文節認識システムを作成した。このシステムは、特定話者の1035単語を含むタスクに対して、第1位で72.0%、第5位までで95.3%の文節認識率を達成した。次に、セパレートベクトル量化および精密な継続時間長制御手法を用いた高精度な音韻モデルを組み込むことにより、システムを改良した。また、ファジィVQコードブックマッピングに基づく話者適応機構も組み入れた。改良したシステムでは、特定話者の場合、文節認識率は第1位で89.5%、第5位までの累積認識率は99.2%（4人の話者の平均）であり、十分な認識性能を達成していると考えられる。話者適応を行なった実験では、第1位で81.6%、第5位までで98.0%の文節認識率を達成した。

第4章では、確率的言語モデルを用いた音声認識システムの認識精度向上について述べる。最初に扱う確率的言語モデルは、日本語の音節の連鎖統計情報、より正確には音節の3つ組確率（trigram）を用いたモデルである。このモデルは、音節列の局所的な誤りを訂正するのに有効である。2番目のモデルは、LR解析アルゴリズムに確率を導入したものである。このモデルでは、まず文脈自由文法の各書き換え規則の適用確率をテキスト・データベースを解析してみることにより求めておく。次に、こうして得られた確率文脈自由文法から、確率付きのLR解析表を作成し、これを解析時に用いる。このモデルは、文法中の各規則がどれだけの頻度で使われるかという統計情報を反映しており、よく使われるような言語表現には高い確率を、まためったに使われないような表現には低い確率を与えるため、音声認識の誤り訂正に有効であると考えられる。3番目の確率的言語モデルは、文脈自由文法

中の書き換え規則の連鎖統計情報を用いたモデルである。文脈自由文法を使って、文を解析すると、その文の解析に使われた書き換え規則の列が得られるが、このモデルはある書き換え規則の後にはどの書き換え規則が使われやすいかという統計情報を反映したモデルである。文脈自由文法では、文脈に依存せずに書き換え規則が適用されるが、このモデルは文脈による書き換え規則の適用の頻度を考慮したモデルとなっており、文脈依存性を確率の形で持っているといえる。以上述べた3つの確率的言語モデルをHMM-LR音声認識システムに組み込み、文節認識実験（1名の話者）で評価した。各モデルを用いることにより、第1位での文節認識率は、88.2%からそれぞれ92.5%、92.1%、91.4%に改善された。また、3つのモデルを同時に用いた場合には、93.2%の認識率を達成することができ、モデルの有効性を示すことができた。

第5章では、2段階LRアルゴリズムを用いた文の認識について述べる。日本語の文は文節列から構成されるため、文節内および文節間の2つのレベルの文法を用いるのが自然である。2段階LRアルゴリズムは、第2章で述べたHMM-LRアルゴリズムの拡張であり、文節間LRバーザによる文節カテゴリ予測と文節内LRバーザによる音韻予測という2段階の予測を用いて、音声認識の探索範囲を縮小することにより、効率よく入力音声を認識する。文節内LRバーザでは、予測された文節カテゴリから、そのカテゴリに属する音韻系列を次々に生成するわけであるが、このために文節内LRバーザは複数の初期状態を持っており、どの文節カテゴリにはどの初期状態が対応するかという表をLR解析表に特化させている。2段階LRアルゴリズムを文節発声による日本語の文の認識に適用した結果、単語認識率95.9%、文認識率84.7%を達成した。一般に、文節間LRバーザは複数の文節カテゴリを予測するため、文節内LRバーザも複数の初期状態を持つことになる。これら複数の状態から同じ音韻が予測された場合、ビームサーチの際に同じ音韻間での競合が起こるために、システムの認識性能を落としてしまうという問題がある。これを解決するために、文節カテゴリごとに初期状態を持たせるのをやめて、LR解析表の各動作項に到達可能なカテゴリを持たせるように改良した。この改良したアルゴリズムを用いることによって、単語認識率を97.5%に、また文認識率を91.2%に改善することができた。

第6章では、音声認識における未知語処理について述べる。通常の音声認識システムでは、システムの語彙中にある単語だけが認識の対象であり、語彙に含まれていない単語、即ち未知語を含む音声を扱うことはできない。しかし、音声認識システムを現実に用いようとした場合、未知語の問題を避けて通ることはできない。ここでは、タスクを記述する通常の文法と、音韻間の制約を記述した文法を統合化することにより、未知語の部分を音韻の列として出力する方法を検討した。また、未知語部分の正しい音韻系列を得るために音節の3つ組確率を同時に用いた。未知語を含む音声の認識実験を行ない、提案した手法が有望なものであることを示した。

第7章では、本研究全体の総括をするとともに、今後の展望について述べている。

付録の2章では、第3章の文節認識実験に用いた文法とその認識結果、および第5章の2段階LRアルゴリズムによる文の認識結果が示されている。