

専攻		学籍番号		指導教官氏名	
申請者氏名	村上 仁一				

論文要旨

論文題目	確率的言語モデルによる自由発話認識に関する研究
------	-------------------------

(要旨 和文 1,200 字程度)

(1)

日本文音声入力においては、音声の持つ物理的特性に着目した音声認識装置の限界を克服するため、日本語の文法や意味を用いた自然言語処理を併用することの必要性が指摘されている。この場合の言語処理の方法として、多くの言語モデルがあるが、大きく分類してルールベースの言語モデルと確率ベースの言語モデルがある。

しかし、言語の確率ベースの研究を行なう場合、基本的には大量のテキストデータ量が必要である。英語ではデータベースの重要性が認識されていて古くから Brown corpus や AP corpus などがあるが、しかし日本語ではコンピュータに読み込める形式で利用できる大量のデータベースが最近まで存在していなかった。そのため、確率的な言語モデルの研究は最近まであまり報告されていなかった。しかし、この状況も新聞記事が CD-ROM で提供されるようになり、国際電気通信基礎技術研究所 (ATR) が各種対話データを販売するなど、状況が変化し始めている。

そこで、本論文では、日本語において N -gram モデルの有効性をシミュレーションや実際の音声認識実験などで定量的に示す。また自由発話認識に向けて、自由発話の言語的特徴や音響的な特徴を研究した。そして実際に自由発話認識にむけた文音声認識のアルゴリズムを提案し、認識実験の結果について報告する。

各章の内容は以下の通りである。

第1章では、本論文の目的、動機について述べた。

第2章では、音声認識システムを実現するために必要な要素技術について述べた。

第3章では言語をマルコフモデルで表現するときのデータ量と収束性について研究した。調査項目としては、主にエントロピーとカバー率である。

第4章では、日本語におけるかなや漢字や品詞の bigram および trigram の有効性を、新聞記事や、医療用 X 線 CT の所見作成や、ATR の国際会議の予約のタスクにおいて、連続分布 HMM と単語 trigram を使用して文認識結果について報告した。

第5章では、自由発話の特徴について言語的な面と音響的な面から研究した。この結果、対話文の 50% は「あー」、「えーと」などの間投詞を含む。また、言い直しは約 10% に出現することが示された。また、4 人の話者について朗読発話と自由発話の音響的な違いについて述べた。そして、音素認識率で両発話の違いを調査したところ、あまり大きな差は無いことが示された。

第6章では、自由発話認識のアルゴリズムとその実験結果について述べた。自由発話では間投詞や、言い淀みや言い誤りおよび言い直しなどが頻繁に出現する。これらの間投詞や言い直しは文の全ての場所に出現する可能性がある。そこでこれらの単語をスキップすることで、自由発話の認識が可能になる。そして実際に実験を行なった結果について述べた。

第7章では、Ergodic HMM を利用した確率付ネットワーク文法の自動学習について述べた。Ergodic HMM と確率つきネットワーク文法が類似した構造を持ち、同種のパラメータで表現される。また、大量のテキストデータを利用してHMM のパラメータを Baum-Welch algorithm で学習できる。実際の会話から作成した単語列を Ergodic HMM に学習させて、確率つきネットワーク文法を自動的に抽出することを試みた。その結果、Ergodic HMM の構造は学習データの特徴をとらえた文法的な特徴を示しており、単語を文中での機能によって分類して出力していることがわかった。さらに、得られた Ergodic HMM を言語モデルとして連続音声認識に用いた。この認識実験の結果、単語 bigram よりも高い性能が得られ、提案したアルゴリズムの有効性が示された。

第8章では、音声情報に含まれている韻律情報の情報量について述べた。韻律情報は F_0 , power, duration などの多くの要素から構成されているが、本章では、この中から特にアクセント句境界の位置およびアクセント核の位置の持つ情報量に焦点を当てて情報量を測定した。実験の結果、アクセント句境界の位置がアクセント情報が持つ情報量は 5.16bit であることが示された。

第9章では、異なる N 個の信号源より生成された信号系列が、どの信号源から生成されたのかを分割・識別する問題について述べた。そして、応用例として複数話者発話の識別をあげ、Ergodic HMM を用いた問題の解決方法を提示した。この実験の結果、複数話者発話の識別においては 341ms 程度の長時間窓分析した LPC ケプストラムを用いることにより、より良好な識別性能が得られること、および尤度の高いモデルを選択することにより平均識別率は向上することが得られた。