

氏名	まさ たま ひろ かず 政 瀧 浩 和
学位(専攻分野)	博 士 (工 学)
学位記番号	論 工 博 第 3441 号
学位授与の日付	平 成 11 年 3 月 23 日
学位授与の要件	学 位 規 則 第 4 条 第 2 項 該 当
学位論文題目	音 声 認 識 ・ 理 解 に お け る 統 計 的 言 語 処 理 の 研 究

(主査)

論文調査委員 教授 堂下修司 教授 石田 亨 教授 美濃導彦

## 論 文 内 容 の 要 旨

連続音声認識において、音素の識別を行う音響モデル単独の性能は、60～70%程度の音素正解率であり、単語間の接続を制限する「言語モデル」や言語理解と組み合わせて認識率を大幅に向上させる手法が必要となる。しかし、非定型性の強い話し言葉にたいして網羅的にかつ矛盾のない文法ルールを書くのは困難であり、近年はテキストデータを統計的に解析して自動的に構築する「統計的言語モデル」と統計的言語理解モデルの成果が期待されている。

本論文は、日本語自然発話の音声認識および音声理解の性能改善を目的として、統計的言語モデル学習用データの整備法、少量の学習データから精度の高い言語モデルの構築法、統計的モデルを用いた頑健な音声理解法などの統計的言語処理モデルの高精度化に関する研究成果をまとめたものであり、序論および結論を含め7章から構成されている。

第1章では、本研究を概観し、研究の背景および目的を述べている。

第2章では、統計処理による音声認識・理解の基本原則を示し、本論文で使用した統計的言語モデル「N-gram」及び「隠れマルコフモデル (HMM)」の動作原理および問題点について説明している。

第3章では、日本語の言語モデルとして、形態素を単位としたN-gramを構築するため、統計モデルによる形態素解析手法を検討し、少量のデータから高い精度が得られる「品詞と可変長形態素列の複合N-gram」を提案している。複合N-gramは、品詞を単位としたN-gramモデルを基本として、特定の形態素は品詞から独立に扱い、さらに特定の形態素は結合して新たな単位として扱い、言語モデルの精度を高めたものである。実験の結果、約40万形態素で学習した複合N-gramは98.6%の形態素解析率が得られ、わずか6千形態素程度の学習でも94.5%という比較的高い精度の形態素解析率が得られた。これらの値は、従来の品詞N-gram、形態素N-gram、およびルールベースの形態素解析よりも高い精度であり、本手法の有効性を確認している。

第4章では、連続音声認識の精度向上を目的として、形態素を単位とした頑健かつ高精度なN-gram言語モデルを提案している。連続音声認識の対象となる話し言葉の言語データを大量に収集するのは困難である。このため、少量の学習データでも頑健でかつ高い予測精度が得られる言語モデルとして、3章の形態素解析でも使用した「品詞および形態素列の複合N-gram」を連続音声認識の言語モデルとして使用することを提案している。連続音声認識の評価実験では、約48万形態素から学習した複合bigramは、パープレキシティが従来の形態素bigramよりも最大約17%低く、また認識誤りは約11%改善され、本モデルの有効性を確認している。

第5章では、目的タスクの言語モデルの精度向上のために、他のタスクのデータで構築したN-gramを目的のタスクに適させる方法として、タスク毎に構築したN-gramの最大事後確率推定 (MAP推定) を用いたタスク適応手法を提案する。MAP推定は、数学的に整備されたパラメータ推定法であり、また形態素組毎に適応の割合を変化させることができるため、従来の単純な線形結合手法よりも精度の向上が望める。約45万形態素、15タスクからなるデータベースを用いた連続音声認識の評価実験の結果、MAP推定によるタスク適応により、適応前よりもパープレキシティが約29%減少、認識誤りは平均

約24%改善することを確認した。また、線形結合タスク適応手法と比較して、本手法の優位性を確認した。

第6章では、音声認識結果の局所的な誤りにたいしても正確かつ頑健な言語理解を得ることを目的として、隠れマルコフモデル (HMM) を用いて、認識結果を中間表現に変換する手法を提案し、出力確率で形態素 (単語) の種類を、状態遷移確率で形態素の並びを確率的に表現する事により、高い理解率を得ている。情報検索システムを対象とした評価実験の結果、約2,600文のデータから学習したモデルは、前後関係のない独立した82文に対して文理解率 (中間表現への変換率) 91.6% が得られ、統計量のみを用いた言語理解法としては良い値を得ている。また、音声認識と言語理解とを合わせた音声理解率は73.2%であったが、これは音声認識の文正解率59.8%よりも約13%高く、若干の認識誤りを含む文からも正しい理解が可能であることを確認している。

第7章では、本論文の結論と、今後の課題及び展望を述べている。

## 論文審査の結果の要旨

計算機と人との自然な対話のための、自由に発話された会話音声の認識と理解システムには、まだ多くの課題が残されている。特に、音響モデルでの音素認識率はまだ不十分で、その誤認識を補うために、精度の高い統計的言語処理が要求される。本論文は、日本語自然発話の音声認識及び理解の精度の向上を目的として、精度の高い統計的言語モデルの学習的構築法と、それによる連続音声認識システムの実現に関する研究の成果をまとめたもので、得られた主な成果は以下のように要約される。

1. 従来からの書き言葉に関する形態素解析とは別に、連続音声に対して統計的処理に基づく形態素解析法として、「品詞、可変長形態素列の複合N-gram」を提案し、実験により比較的少量のコーパスデータから精度のよいN-gramが得られることを示した。
2. 品詞、可変長形態素列の複合N-gramを言語モデルとして用いて、連続音声認識の実験を行い、評価実験により、従来の方法に比べて、より小さいパープレキシティと誤認識率が得られることを示した。
3. タスクを設定する事により、理論的には認識のスコアが向上するが、実際にはタスク毎のN-gram構成のためのコーパスサイズが小さくなることから、N-gramの精度が下がり、認識率を悪くする。これに対して、目的のタスクのN-gramをベースとして、他の多くのタスクのN-gramを目的タスクに適応させる事により十分の量と精度のN-gramを得る方法を考案し、それを最大事後確率推定法により実現し、有効性を示した。
4. 音声認識レベルと言語理解・レベルを統計的手法により包括するモデルとして、マルコフモデルを採用し、音声認識誤り等による非文法的な文に対しても一定のロバスト性が得られることを示した。

以上要するに、本論文は、日本語の自然発話に対して、精度の高い統計的言語モデルの学習的構築法と、それによる高い精度の連続音声認識システムの構成法を示し、実験的にその有効性を明らかにしたものであり、得られた成果は實際上、学術上寄与するところが少なくない。よって本論文は博士 (工学) の学位論文として価値あるものと認める。また、平成11年2月9日、論文内容とそれに関連した試問を行った結果合格と認めた。