

外95-37

早稲田大学大学院理工学研究科

博士論文概要

論文題目

者声認誠における
識別學習理論に基づくスボッタ設計法
スボッティング"に関する研究

申請者

小森 隆

Takashi KOMORI

1995年12月

概 要

電子計算機の普及が進むにつれ、ヒューマンインターフェースの重要性が増している。なかでも人間が発話した音声を計算機に自動認識させる音声認識の技術は早くから関心を集めしており、長い研究の歴史がある。

初期の音声認識研究において成果をあげたのは、非線形な時間伸縮とともに時系列パターン照合を動的計画法に基づいて効率よく行なう計算法であった。このアルゴリズムは動的時間伸縮 (dynamic time warping; DTW) 法あるいは DP マッチング法と呼ばれ、離散単語認識や連続単語認識などの限定された音声認識課題に適用された。しかし DTW 法には計算量が多いという問題があった。また、単語ごとに参照パターンを用意する必要があるため語彙の拡張や変更に対応しにくいという問題があった。特に、不特定話者の音声を認識しようとした場合にはこの問題はさらに深刻であった。単語単位の参照パターンでは隣り合う単語間の調音結合にうまく対処できないという問題もあった。

その後、不特定話者の大語彙単語認識に適するアルゴリズムとして隠れマルコフモデル (hidden Markov model; HMM) を用いたものが提案され、音声認識技術の主流となった。HMM 法では、音素やそれに類する短い音声基本単位を、確率分布モデルで表現する。このため、多数の話者の音声の音響的多様性を少ないパラメータ数で表現できる。また、音響モデルを連結することでさらに長い音声単位のモデルを動的に構成できる。こうした利点から、HMM 法は文法的な制約条件と組み合わされ、文章音声認識にも適用された。とはいっても、大語彙の単語を含む文章音声の認識を試みる場合、文候補の数は組み合せ爆発的に大きくなるため、計算量の多さは依然問題であり、ビームサーチ法などのヒューリスティックな計算省略法を用いざるを得ない。特に、任意発話を認識しようとした場合には、受理すべき発話内容が非常に広範囲に渡り、妥当な制約文法を用意することさえ困難になる。

こうした背景から、キーワードスポットティングに基づく接近法が研究の関心を呼んでいる。連続音声発話中から認識の鍵となる単語をその時刻とともに検出するキーワードスポットティングは、任意発話音声の理解のための有望な枠組であると考えられ、実際、古くから音声理解の実験システムなどに用いられてきている。しかしスポットティングには、単純な分類問題とは異なる固有の問題点が存在し、それらに対する明確な解は未だ見出されていない。まず、単語等の小さい言語単位で検出の決定が行なわれるため妥当な文法的制約を導入することが難しく、音響的な情報のみに頼って決定を行わなければならない。また、音声発話に含まれるキーワードをできるだけもらさず検出するために検出の条件を緩めると、余分なキーワード候補 (付加誤り) が大量に発生してしまい、やはり最終的な誤認識の原因となる。さらに、任意発話では音響的変形も大きいため、そのことも現実的に認識を難しくして

いる。したがって、スポットティングを行なうシステムには、音響的な情報のみに基づいて精度の高い認識を行うことが求められる。その実現のためには、理論的な裏付けを持ち、実際の任意発話音声を学習データとして直接用いるスポット設計法が必要である。

この観点から、本研究では、スポットティングにおける認識誤りを最小化するための新しい統計的スポット設計法 (あるいはスポット学習法) を提案する。本研究では、音声の発生の仕組みが未知であることからノンパラメトリックな方法が望ましいと考え、Katagiri らによつて提案された一般化確率的降下法を基本的概念として用いることにした。一般化確率的降下法は、分類等の処理における誤りの個数を損失関数に直接反映させ、最小化するための学習法を与えるものである。本研究ではまず、音声認識の古典的枠組みである DTW 法による分類器の設計に一般化確率的降下法を適用し、その評価を行なう。次に、単一の単語スポットにおいて一般化確率的降下法に基づく設計法を提案する。さらに、複数のスポットティング結果に基づいた統合的決定における誤りを最小化するための学習法を、やはり一般化確率的降下法に基づいて提案する。

本論文は以下の 5 章により構成される。

第 1 章では、音声認識研究、特にキーワードスポットティングの研究における現状と問題点を概観し、本研究の課題を明らかにした。

第 2 章では、一般化確率的降下法に基づき、DTW 法に基づく音声パターン分類器のための具体的な最小分類誤り学習法を導き出した。この章ではまず、音声パターン分類器が、最大尤度規準や最小歪規準などの従来の設計規準ではなく、最小分類誤りという直接的な設計規準を用いて設計されるべきであることを述べた。次に、DTW 法を用いた場合の音声パターン分類問題を厳密に定式化し、最小分類誤り規準に基づく学習法を一般化確率的降下法にしたがって実際に導き出した。G-訓練則と名付けられたこの学習法は計算量の点で問題があるため、これをさらに簡便化し、より実地的な学習法である S-訓練則を導出した。S-訓練則については、実際の音声パターンの認識課題 2 種類を用いて評価実験を実施した。実験の結果、本章で提案した学習アルゴリズムが従来の DTW 法による音声分類器の実際の音声パターンに対する識別能力の増大に貢献することが実証された。また、パターン分類のための代表的な識別学習アルゴリズムとして学習ベクトル量子化 (learning vector quantization; LVQ) が知られているが、S-訓練則がその一般形と見なされることを述べ、この章で提案したアルゴリズムがパターン分類問題一般に応用可能であることを示した。

第 3 章では、スポットティング誤りの数を最小化することを設計目標とするスポット設計法を提案した。この章では、DTW 法に基づく距離としきい値の比較によってスポットティング決定を行なうスポットを例として取り上げ、スポットのパラメータとスポットティング誤りとの関係を定式化した。さらに、具体的な最小スポットティング誤り学習法を、第 2 章と同様

に一般化確率的降下法に基づき導き出した。この章において提案した学習法は、脱落誤りと付加誤りに異なる重みを持たせ、要求に応じて柔軟な最適化を実行することができる。定式化の2種類の様式について、日本語子音のスポットティング実験によって評価した。学習実験の結果、この章で提案する学習法の非常に高い利用可能性が示された。この章では、スポットティング誤りの最小という新しい設計目標によるパラメータの学習を行うことにより、スポットタの構成を変えることなくその性能を向上させることができることが確認された。

第4章では、個々のキーワードスポットティングの結果から得られるキーワード列レベルでの認識誤りを最小化するための学習法を、やはり一般化確率的降下法に基づき導き出した。まず、この課題が分類課題の一種であることを示し、スポットティングのための新たな判別関数を厳密な定式化に基づいて提案した。定式化の過程で、個々のキーワードスポットティングを組み合わせてキーワード列を認識する場合、スコアとしては対数事後オッズの推定値を用いることが妥当であり、従来用いられていたヒューリスティックなスコア正規化法では不十分であることを示した。次に、第2章と同様、一般化確率的降下法に基づいて、その定式化に基づく具体的な訓練則を導き出した。また、認識誤りと計算量という相反する設計目標についてスポットタのパラメータを一貫的に最適化する方法についても提案した。さらに、実際の連続音声中からのキーワードスポットティング課題において評価実験を実施した結果、本章で提案した学習法の有用性が明らかにされた。

最後に、第5章では、本研究を総括し、今後の研究の展望について述べた。

本研究において提案した手法は、いずれも基本的には特定の認識器構造を前提としているため、どのような認識器構造に対しても本質的な変更を伴わずに適用することが可能である。とはいっても、実験による評価は特定の認識器に限らざるを得なかった。したがって、さまざまな認識器構造におけるこれらの手法の実地的評価が今後の研究課題として残されている。また、本研究において用いた一般化確率的降下法は、勾配降下型アルゴリズムであるため、保証されているのは局所的最適解への収束のみである。擬似焼きなまし法の適用などによる局所的最適解からの脱出を図るために工夫も将来的な研究課題であると考えられる。