

TR-IT-0025

連続音声認識の研究
A Study on Continuous Speech Recognition

山口 耕市
Kouichi YAMAGUCHI

1993年10月

概要

著者が1991年5月1日から1993年10月31日まで、ATR自動翻訳電話研究所及びATR音声翻訳通信研究所で行なった、連続音声認識における探索問題を中心とした諸研究について報告する。研究内容は主として次の4つである。

- 混合連続分布型 HMM-LR
- A* サーチアルゴリズムを用いた HMM-LR
- 学習可能な適応的探索手法
- 話者一貫性パーズングを用いた不特定話者連続音声認識

これらに関連する研究発表論文を添付した。

目次

1 研究の概要	1
2 発表論文リスト	2
3 付録	3
A 付録: ICSLP92	4
B 付録: 日本音響学会平成 4 年度秋季	8
C 付録: ICASSP93	10
D 付録: 日本音響学会平成 5 年度秋季	14

1 研究の概要

著者が1991年5月1日から1993年10月31日まで、ATR自動翻訳電話研究所及びATR音声翻訳通信研究所で行なった、連続音声認識における探索問題を中心にした諸研究について報告する。今までの主な研究の概要を以下に示す。

- 混合連続分布型 HMM-LR 開発

従来音素モデルとして離散分布型 HMM を用いていた HMM-LR 文節音声認識システムに対し、より精度の高い認識系をつくることのできる混合連続分布型 HMM に置き換えた。これらの改良により、多大な計算を要していた継続時間制御を用いなくても、従来の離散分布型 HMM-LR を若干上回る性能が得られた。また、HMM の学習サンプルに文節発声データを追加することによって文節認識性能が飛躍的に向上することを見出し、発話様式への適応の重要性を示した。

- A* サーチアルゴリズムの導入

HMM-LR 文節音声認識システムの高速度の一環として、また文音声認識への発展を考慮し、サーチ方法を改良した。従来から用いているビームサーチ法では最適解が枝刈りされてしまふ一方、枝刈りを避けるためにビーム幅を大きく設定すると計算量が増える。そこで、最適解が保証されていて、かつ高速なサーチが可能な A* アルゴリズムを導入した。認識性能は若干低下するものの、探索空間を平均で約 1/3 に、サンプルによっては 1/10 以下に削減できることを示した。

- 学習可能な適応的探索手法の検討

HMM-LR 連続音声認識システムの高速度の一環として、HMM-LR 文節音声認識システムの高速度をめざし、ニューラルネットワークによる適応的ビームサーチを導入した。これは学習可能なサーチ方法という点で、探索問題における新しいパラダイムを提案した。

- 話者一貫性パーズングの導入

話者一貫性原理に基づいた不特定話者連続音声認識方式である「話者一貫性パーズング」を提案した。着眼点は、不特定話者の発声であっても、話者は発声を通して同一である点である。そこで、連続音声認識のパーザにおいて、発声内容と話者タイプの2方向を探索の対象とする。認識動作完了とともに、以後の話者適応で使うべき標準話者のクラスタが選出されるという利点を持っている。つまり動的な話者適応の一つの実現手段にもなりえる。認識性能に関しては、従来の不特定話者音声認識システムが抱えていた、広がり過ぎた混合分布による認識性能劣化を防ぐ効果を持つ。また、順位の低い仮説しかもたない話者は枝刈りされ、その時点から以後その話者の尤度は計算する必要がなくなることにより、HMM のフレーム尤度計算量の削減にも貢献する。

SSS-LR 連続音声認識システム上で不特定話者音声認識実験を行ない、従来法の不特定話者法との認識性能を比較した。本方式は標準話者を3クラスタに選択した場合、不特定話者法の認識率を上回る性能を得た。

- その他の成果

- ・ 日本語の鼻音に関して調音点による分類 $[m, n, ng]$ と音素による分類 $/m, n, N/$ との識別性能比較実験を行ない、概ね $[m, n, ng]$ より $/m, n, N/$ のほうが性能が高いことを見出し、音素モデル作成に指針を与えた。

- ・ 離散 HMM で動作していた HMM-LR 文認識を、混合連続分布型 HMM に置き換え、種々の問題点を検討した。文認識では文節認識に比べ、仮説の可能性が増大するため探索空間が大きくなり、効率良いサーチ手法の必要性を確認した。

2 発表論文リスト

1. 山口耕市, 嵯峨山茂樹: “混合連続分布型 HMM を用いた HMM-LR 連続音声認識,” 日本音響学会平成 4 年度春季研究発表会講演論文集, 1-P-5, pp. 113-114, (1992.03).
2. 山口耕市, 嵯峨山茂樹, 北研二, Frank K. Soong: “HMM-LR 連続音声認識における A* アルゴリズムを用いた探索手法の検討,” 日本音響学会平成 4 年度春季研究発表会講演論文集, 3-1-9, pp. 87-88, (1992.03).
3. Kouichi Yamaguchi, Shigeki Sagayama, Kenji Kita, Frank K. Soong: “Continuous Mixture HMM-LR Using the A* Algorithm for Continuous Speech Recognition,” Proc. of 1992 International Conference on Spoken Language Processing, We.sAM.1.2, pp. 301-304, (1992.10).
4. 山口耕市, 嵯峨山茂樹: “ニューラルネットワークを用いた適応的ビームサーチによる HMM-LR 連続音声認識,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 3-7-2, pp. 165-166 (1992.10).
5. 山口耕市, 永井明人, 鷹見淳一, 大倉計美, 小坂哲夫, 福沢圭二, 加藤喜永, 北研二, Harald Singer, 村上仁一, 杉山雅英, 嵯峨山茂樹, 服部浩明, 小森康弘, 沢井秀文, 花沢利行, 中村哲, 甲斐充彦, 南泰浩, 川端豪, 鹿野清宏: “ATREUS: ATR における連続音声認識諸方式の比較,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 2-Q-5, pp. 181-182 (1992.10).
6. S. Sagayama, M. Sugiyama, K. Ohkura, J. Takami, A. Nagai, H. Singer, H. Hattori, K. Fukuzawa, Y. Kato, K. Yamaguchi, J. Murakami, and A. Kurematsu: “ATREUS: Continuous Speech Recognition Systems at ATR Interpreting Telephony Research Laboratories,” Proc. SST92 (Australia) (1992.12).
7. 山口耕市, 嵯峨山茂樹: “ニューラルネットワークによる学習可能な探索手法を用いた連続音声認識,” 信学技報, SP92-123, pp. 1-7, (1993.1).
8. 永井明人, 山口耕市, 鷹見淳一, 大倉計美, 小坂哲夫, 福沢圭一, 加藤喜永, Harald Singer, 村上仁一, 杉山雅英, 嵯峨山茂樹, 保坂順子, 森元逞, 北研二 (徳島大学), 服部浩明 (日本電気), 小森康弘 (キヤノン), 沢井秀文 (リコー), 花沢利行 (三菱電機), 中村哲 (シャープ), 甲斐充彦 (豊橋技科大), 南泰浩, 川端豪, 鹿野清宏 (以上 NTT), 樽松明, “ATR における連続音声認識システム ‘ATREUS’ の諸方式と性能,” 信学技報, SP92-122, pp. 51-58, (1993.1).
9. Kouichi Yamaguchi and Sigeki Sagayama: “A Neural Network Controlled Adaptive Search Strategy for HMM-based Speech Recognition,” Proc. ICASSP93 (Minneapolis, USA), vol. II, pp. 582-585, (1993.4).
10. A. Nagai, K. Yamaguchi and A. Kurematsu, “ATREUS : A Comparative Study of Continuous Speech Recognition Systems at ATR,” Proc. ICASSP93 (Minneapolis, USA), vol. II, pp. 139-142, (1993.4).
11. 山口耕市, Harald Singer, 松永昭一, 嵯峨山茂樹: “話者一貫性パーズングを用いた不特定話者連続音声認識,” 日本音響学会平成 5 年度秋季研究発表会講演論文集, 2-7-4, pp. 77-78 (1993.10).

12. Kouichi Yamaguchi, Harald Singer and Sigeki Sagayama: "Speaker-Consistent Parsing for Speaker-Independent Continuous Speech Recognition," Proc. ICASSP94 (Adelaide, Australia), 投稿中 (1994.4).

注: 文献略語

- 信学技報: 電子情報通信学会技術報告
- ICASSP: International Conference on Acoustics, Speech, and Signal Processing
- ICSLP: International Conference on Spoken Language Processing
- SST: international conference on Speech Science and Technology

3 付録

これらの方法、実験結果を付録に示す。