様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月 １６ 日現在

機関番号：１２６０８
研究種目：挑戦的研究（萌芽）
研究期間：2017～2019
課題番号：１７Ｋ２０００１

研究課題名（和文）全ベイズモデルに基づく音声認識システム学習のデータ無制約化

研究課題名（英文）Constraint Free Training of Speech Recognition Systems Based on Full Bayes Modeling

研究代表者
篠崎　隆宏（Shinozaki, Takahiro）

東京工業大学・工学院・准教授

研究者番号：８０４４７９０３

交付決定額（研究期間全体）：（直接経費）　　4,800,000円

研究成果の概要（和文）：音声認識を様々なタスクにおいて実用的なものとするためには、認識システムの学習において教師あり学習への依存度を減らし、システムをより自律的なものへとする必要がある。本研究では、ノンパラメトリックベイズ法と重み付き有限トランスデューサ技術を応用し、対応の無い音素データとテキストデータから、自動的に発音辞書を拡張する手法を提案した。また書き起こしテキストを用いずにEncoder-Decoder型の音声認識システム全体を方策関数として認識結果のスカラー評価値をもとにシステムを強化学習する方法について取り組み、学習効率を大幅に向上させる手法を提案した。

研究成果の学術的意義や社会的意義
人間は成長の過程でほとんど無意識のうちに平均して一日５単語以上を学習する優れた言語学習能力を持っている。それに対して現在の音声認識システムは教師あり学習に頼っておりシステム開発に多大な手間を必要とするとともに、日々生み出される新しい単語や小さなコミュニティ内でのみ使用される表現などを自動的に学習する能力を欠いている問題がある。人と機械の間での自然な音声対話の実現を目指し、本研究では自律的な学習技術の実現に取り組んだ。従来の教師あり学習に代わる教師なし学習や強化学習による学習手法を提案し、実験により有効性を示した。

研究成果の概要（英文）：The dependency on supervised learning using paired data is a major bottle-neck of current speech recognition systems. The goal of this research is to improve the flexibility of the system learning by using unpaired data. We have proposed a method to automatically extend the pronunciation dictionary from unmatched phoneme data and text data by applying the nonparametric Bayes method and weighted finite transducer. We have also worked on reinforcement learning of speech recognition systems by formulating the whole encoder-decoder based system as a policy function. We have shown that our proposed reinforcement learning methods significantly improve learning efficiency.

研究分野：音声言語情報処理

キーワード：音声認識　教師なし学習　半教師あり学習　強化学習　ノンパラメトリックベイズ法

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

深層学習の進展などに伴い音声認識システムの性能が向上し、スマートフォンなどを用いたアプリケーションに実用されつつある。しかし、現在の音声認識システムの性能は大量の書き起こしテキスト付き音声データを用いた教師あり学習に依存しており、また認識タスクの違いにも脆弱な問題がある。音声の書き起こしを行う作業は非常に手間のかかる作業であり、大きな労力と費用がかかる。このため、実際に高い認識性能を実現できているのは一部の言語の一部のタスクに限られている。また、あるタスクにおいて一度高性能なシステムを構築したとしても、新しい言葉が日々登場するなど言語は常に変化しているため、認識性能を維持するためには人手による継続的なメンテナンスが必要となる問題がある。

一方、人において音声とそのテキスト表現が同時に与えられる教師付き学習は、家庭や学校などで読み書きを教わる状況に相当する。これにより効率的な学習が可能であるが、日常的な生活を通して日々行われている言語学習を含めた全体から見れば学習の初期における比較的限定された場面である。多くの場合は対話や読書などの形で音声あるいはテキストのみのデータに接する状況で学習が行われている。対象とするテキストも通常は形態素解析がされたものではないため単語境界は与えられず、また読みの不明な単語が出現する場合もしばしばである。それでも言語のコンテキスト情報などをもとに学習を進めている。

すなわち現在の音声認識技術は認識精度についてはタスクによっては人に近い性能が得られるようになりつつあるものの、学習能力の点では人に大きく劣っている問題がある。音声認識を幅広い応用分野において有効なものとするためには、柔軟で自律的な学習技術の実現が必要となる。

２．研究の目的

音声認識システムが人と同様に、与えられた種々のデータを自由に学習に用いることを可能にし、それによりシステム開発時の認識タスクの想定に制約されることなく、変化する環境に対して継続的に性能を向上させられる学習の仕組みを実現することを目的としている。具体的には、通常の書き起こしテキスト付き音声データだけでなく、対となるテキストの存在しない音声データや発音が不明な単語を含むテキストデータなど、入手できる様々なデータを効果的に使用できる学習アルゴリズムの実現に取り組む。

３．研究の方法

音声認識システムは音響モデルや言語モデル、発音辞書などから構成される。これらは従来、それぞれ個別に学習あるいは作成されてきた。しかし例えば、表記は既知であるが発音が未知の単語の発音をその単語の発声を含む音声データから学習するためには、まず発音未知の単語が音声に含まれていることを認識する必要がある。そしてさらに、それがどの表記の単語に対応するのかを言語的なコンテキストから推定する必要がある。そのためには、既知の音響的および言語的知識を総合的に活用することが必要となる。書き起こしテキストのない音声データからの音響モデルや言語モデルの学習も同様に、総合的な知識の活用が重要となる。そこで、音声認識システムとその学習プロセス全体を統一されたベイズモデルとして定式化し、そのモデル上でデータに応じたベイズ推論を行うことで、全体を考慮しながら効果的な学習を行うアプローチを提案する。発音辞書についてもノンパラメトリックベイズ法を応用して確率変数として扱うことで、発音が未知の単語や、単語自体が未知の単語をデータから学習することが可能となる。

また、Encoder-Decoder 型の音声認識システム全体を方策関数として、書き起こしテキストを用いずに認識結果のスカラー評価値をもとにシステムを強化学習する方法について検討および実験を進める。これは、クラウド上に構築された音声認識サーバーが多数のユーザーに認識サービスを提供する際

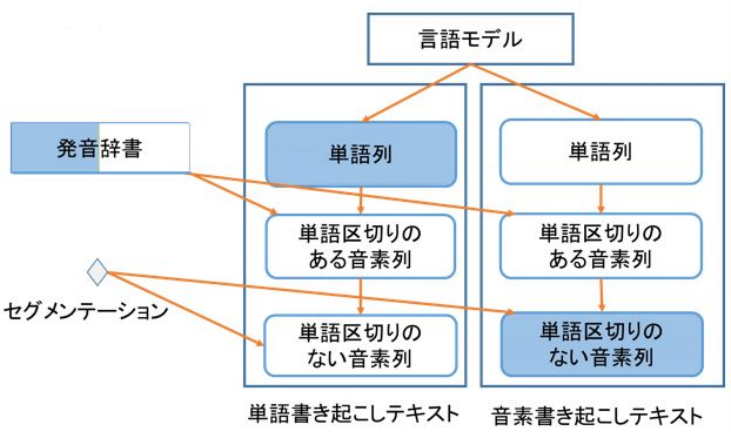

図 １半教師あり発音辞書学習のためのベイズモデル

に、ユーザーからのわずかなフィードバックを大量に集めることでシステム性能を向上させる応用を想定したものである。

４．研究成果

Reference	you’re a friend in bad times as well as good
Epoch1	you’re a friend in ,comma as well as good
Epoch2	you’re a friend in bad time close as well as good
Epoch3	you’re a friend in bad times as well as good

図 2　発音辞書学習による単語認識率の改善例。Reference が正解文、Epoch が学習の繰り返し数を示す

【発音辞書学習】
音声認識システムにおける発音辞書は、単語とその発音の対から構成される。発音辞書によって、各単語についてその発音に対応した音素列が与えられる。しばしば、出現頻度に応じた確率重みを持たせることで一つの単語が複数の発音を持つことができるように発音辞書は設計される。確率モデルとして見ると、これは各単語において混合要素が発音、混合重みが確率として構成された有限混合モデルとして考えることができる。

発音辞書の半教師あり学習では、発音未知の単語の発音の候補として様々な可能性を考慮する必要がある。そこで第一に、各単語にどのような発音でも割り当てられるように有限混合モデルから無限混合モデルへの拡張を行った。具体的には、各単語の発音を、音素列を要素とする無限の要素を持つ離散分布として定式化した。発音分布の事前分布は、音素列を生成する分布を基底分布とするディリクレ過程により与える。学習データとして音素列と単語列の対が与えられる場合は、発音辞書に基づいた発音の予測分布は中華料理店過程（CRP）により求められる。音素列は音声からの音素を認識することで得られる。しかし音素列と単語列が対となっていないより一般的な状況において発音辞書の学習を行うためには、文脈を手掛かりとしたより高度な推論が必要となる。そこで第二に、単語列に関する言語的情報、音声認識仮説における音素列のパターン情報、および単語発音に関する部分的な既存の知識をつなぎあわせて半教師あり発音辞書学習を行うベイズモデルの提案を行った。モデルの構成を図１に示す。提案ベイズモデルは、発音辞書、言語モデル、および単語セグメンテーションモデルから構成される。このうち、発音辞書と言語モデルが学習対象であり、言語モデルには階層 Pitman-Yor 過程を用いている。

提案する発音辞書学習のためのベイズモデルはコンテキスト情報の利用のために言語モデルを含み、音素及び単語の時系列を扱うために大規模で複雑な構成となる。そこで各要素を統合しながら確率推論を行う仕組みとして、重み付き有限状態トランスデューサー(WFST)を応用した。WFST として実装した各要素を合成演算により統合し全体を実装するアプローチであるが、一般的な合成演算を用いると本モデルにおいて必要となる中間段階の確率変数が消去されてしまう問題がある。そこで、合成の際に中間の確率変数の値を合成後の WFST の入力シンボルに繰り込むことで保持する手法を提案し実装した。この枠組みのもと、ギブスサンプリングにより確率推論を行う。

実験は英語および日本語データを用いて行った。発音辞書中で一定割合の単語の発音が未知の状態において、システムには対となっていない独立な音素列と単語列の集合が学習データとして与えられる。音素列と単語列は同一言語からサンプルされたものであり、個別の対応はなくても全体として一つの言語における発話の分布に従っている。この条件において、単語出現のコンテキストを手掛かりに単語の発音を学習することがタスクである。実験に用いたどちらの言語においても、学習を進めることで単語発音の学習が進み、単語誤り率が減少することを示した。図２は結果の一例であり、当初は間違った単語が認識されていた箇所が学習の進展とともに正しく認識される様子を示している。

【対話に基づく音声モデル学習】
同じ言語からサンプルされた対となっていない音声及びテキストの集合を用いる教師なし学習や半教師あり学習に加えて、システムがユーザーとの対話の中で何らかのフィードバックを受け取りながら性能を改善していく学習が考えられる。ここでフィードバックとは、いくつかの認識結果候補の中からユーザーが正しいものを選択することや、認識結果の品質に対して評価の数値を与えることである。そのような学習を可能にするため、音声認識システムの強化学習やブラックボックス最適化について取り組みを行った。End-to-End 音声認識システムは入力音声が与えられた条件で単語列の事後確率を計算しており、全体を方策関数ととらえることができる。強化学習に適したモデル構造や目的関数を提案し、学習初期段階のようにシステムの認識性能が低い場合でも学習が進む手法を提案した。さらに提案法では、フィードバックとなる評価スコアにノイズが含まれている場合でもシステムの学習が可能であることを示した。

５．主な発表論文等

〔雑誌論文〕　計15件（うち査読付論文　9件／うち国際共著　4件／うちオープンアクセス　5件）

１．著者名 Takafumi Moriya, Tomohiro Tanaka, Takahiro Shinozaki, Shinji Watanabe, Kevin Duh	４．巻 27
２．論文標題 Evolution-Strategy-Based Automation of System Development for High-Performance Speech Recognition	５．発行年 2018年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 77-88
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2018.2871755	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Xu Han, Takahiro Shinozaki, Ryota Kobayashi	４．巻 -
２．論文標題 Effective and Stable Neuron Model Optimization Based on Aggregated CMA-ES	５．発行年 2019年
３．雑誌名 Proc. IEEE ICASSP	６．最初と最後の頁 1264-1268
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Bairong Zhuang, Wenbo Wang, Takahiro Shinozaki	４．巻 -
２．論文標題 Investigation of Attention-Based Multimodal Fusion and Maximum Mutual Information Objective for DSTC7 Track3	５．発行年 2019年
３．雑誌名 Proc. DSTC7	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Yilong Peng, Hayato Shibata, Takahiro Shinozaki	４．巻 -
２．論文標題 Reward Only Training of Encoder-Decoder Digit Recognition Systems Based on Policy Gradient Methods	５．発行年 2018年
３．雑誌名 Proc. APSIPA	６．最初と最後の頁 1934-1939
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Tomohiro Tanaka, Takahiro Shinozaki	4．巻 -
2．論文標題 F-Measure Based End-To-End Optimization of Neural Network Keyword Detectors	5．発行年 2018年
3．雑誌名 Proc. APSIPA	6．最初と最後の頁 1456-1461
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Taku Kato, Takahiro Shinozaki	4．巻 -
2．論文標題 Reinforcement Learning of Speech Recognition System Based on Policy Gradient and Hypothesis Selection	5．発行年 2018年
3．雑誌名 Proc. IEEE ICASSP	6．最初と最後の頁 5759-5763
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ICASSP.2018.8462656	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 篠崎　隆宏	4．巻 67
2．論文標題 自動音声認識技術と英語教育--仕組みと研究動向，今できること・できないこと--	5．発行年 2019年
3．雑誌名 英語教育	6．最初と最後の頁 40-41
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Hao Qin, Takahiro Shinozaki, Kevin Duh	4．巻 -
2．論文標題 Evolution Strategy Based Automatic Tuning of Neural Machine Translation Systems	5．発行年 2017年
3．雑誌名 Proc. International Workshop on Spoken Language Translation (IWSLT)	6．最初と最後の頁 120～128
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Zhuang Bairong, Wang Wenbo, Li Zhiyu, Zheng Chonghui, Takahiro Shinozaki	４．巻 -
２．論文標題 Comparative Analysis of Word Embedding Methods for DSTC6 End-to-End Conversation Modeling Track[C]	５．発行年 2017年
３．雑誌名 Proc. Dialog System Technology Challenges (DSTC6)	６．最初と最後の頁 1～5
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 池下裕紀，篠崎隆宏	４．巻 -
２．論文標題 音声認識仮説を用いたベイズ的半教師あり発音辞書学習の検討	５．発行年 2018年
３．雑誌名 日本音響学会2018年春季研究発表会講演論文集	６．最初と最後の頁 123～124
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 加藤拓，篠崎隆宏	４．巻 -
２．論文標題 方策勾配法と仮説選択に基づくDNN音声認識システムの強化学習	５．発行年 2018年
３．雑誌名 日本音響学会2018年春季研究発表会講演論文集	６．最初と最後の頁 15～16
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 鄭 崇輝，李 知雨，王 文博，庄 佰融，篠崎 隆宏	４．巻 -
２．論文標題 End-to-Endニューラル対話モデルにおける単語分散表現の比較検討	５．発行年 2018年
３．雑誌名 日本音響学会2018年春季研究発表会講演論文集	６．最初と最後の頁 125～126
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 加藤 拓，篠崎 隆宏	４．巻 Vol.2017-SLP-119
２．論文標題 英語学習者の発声自動評価を目的としたDNN音声認識システムの検討	５．発行年 2017年
３．雑誌名 情報処理学会研究報告	６．最初と最後の頁 1～4
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 池下裕紀，篠崎隆宏，渡部晋治，持橋大地，Graham Neubig	４．巻 Vol.2017-SLP-118
２．論文標題 ベイズ推論を用いた半教師あり学習の日本語適用	５．発行年 2017年
３．雑誌名 情報処理学会研究報告	６．最初と最後の頁 1～4
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Takahiro Shinozaki, Shinji Watanabe, Daichi Mochihashi, Graham Neubig	４．巻 -
２．論文標題 Semi-Supervised Learning of a Pronunciation Dictionary from Disjoint Phonemic Transcripts and Text	５．発行年 2017年
３．雑誌名 Proc. Interspeech	６．最初と最後の頁 2546-2550
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕　計18件（うち招待講演　0件／うち国際学会　2件）

１．発表者名 田中智宏，篠崎隆宏
２．発表標題 連続単語検出のための 2D-RNN を用いた End-to-EndDPマッチング
３．学会等名 日本音響学会　2019年　春季研究発表会
４．発表年 2019年

１．発表者名
田中智宏，篠崎隆宏

２．発表標題
連続対応検出ネットワークによる音声動画からの教師なし物体セグメンテーションおよび関連学習の検討

３．学会等名
日本音響学会　2019年　春季研究発表会

４．発表年
2019年

１．発表者名
PengYilong，篠崎隆宏

２．発表標題
大規模 End-to-End 音声認識システムの教師なし強化学習の実現に向けた検討

３．学会等名
日本音響学会　2019年　春季研究発表会

４．発表年
2019年

１．発表者名
王 文博，庄 佰融，篠崎 隆宏

２．発表標題
Analysis of Attention-Based Multimodal Fusion and Maximum Mutual Information Objective for DSTC7 Audio Visual Scene-Aware Dialog Track

３．学会等名
日本音響学会　2019年　春季研究発表会

４．発表年
2019年

１．発表者名
Yi Liu, Takahiro Shinozaki

２．発表標題
I-vector Domain Adaptation Using Cycle-Consistent Adversarial Networks for Speaker Recognition

３．学会等名
情報処理学会 SLP-126

４．発表年
2019年

１．発表者名
田中智宏，篠崎隆宏

２．発表標題
マルチゲートGRUユニットを用いた2D-RNNによるEnd-to-End始終端フリー単語検出

３．学会等名
情報処理学会 SLP-125

４．発表年
2018年

１．発表者名
Wenbo Wang, Bairong Zhuang, Takahiro Shinozaki

２．発表標題
Improving the audio visual scene-aware dialog system in DSTC7 by using attentional multimodal fusion and MMI objective

３．学会等名
情報処理学会 SLP-125

４．発表年
2018年

１．発表者名
PengYilong，柴田駿人，篠崎隆宏

２．発表標題
音声認識システムの教師なし強化学習における報酬と報酬ノイズの影響の検討

３．学会等名
日本音響学会　2018年　秋季研究発表会

４．発表年
2018年

１．発表者名
田中智宏，篠崎隆宏

２．発表標題
単語検出性能を目的関数とした単語検出器学習法の提案

３．学会等名
日本音響学会　2018年　秋季研究発表会

４．発表年
2018年

１．発表者名
柴田駿人，PengYilong，篠崎隆宏

２．発表標題
強化学習による報酬のみを用いたend-to-end 認識システム学習

３．学会等名
日本音響学会　2018年　秋季研究発表会

４．発表年
2018年

１．発表者名
PengYilong，柴田駿人，篠崎隆宏

２．発表標題
End-to-end音声認識システムの強化学習の検討

３．学会等名
情報処理学会 SLP-123

４．発表年
2018年

１．発表者名
Hao Qin

２．発表標題
Evolution Strategy Based Automatic Tuning of Neural Machine Translation Systems

３．学会等名
International Workshop on Spoken Language Translation（国際学会）

４．発表年
2017年

１．発表者名
Zhuang Bairong

２．発表標題
Comparative Analysis of Word Embedding Methods for DSTC6 End-to-End Conversation Modeling Track[C]

３．学会等名
Dialog System Technology Challenges (DSTC6)（国際学会）

４．発表年
2017年

１．発表者名
池下 裕紀

２．発表標題
音声認識仮説を用いたベイズ的半教師あり発音辞書学習の検討

３．学会等名
日本音響学会春季研究発表会

４．発表年
2018年

１．発表者名
加藤 拓

２．発表標題
方策勾配法と仮説選択に基づくDNN音声認識システムの強化学習

３．学会等名
日本音響学会春季研究発表会

４．発表年
2018年

１．発表者名
鄭 崇輝

２．発表標題
End-to-Endニューラル対話モデルにおける単語分散表現の比較検討

３．学会等名
日本音響学会春季研究発表会

４．発表年
2018年

１．発表者名
加藤 拓

２．発表標題
英語学習者の発声自動評価を目的としたDNN音声認識システムの検討

３．学会等名
情報処理学会音声言語情報処理研究会

４．発表年
2017年

１．発表者名
池下 裕紀

２．発表標題
ベイズ推論を用いた半教師あり学習の日本語適用

３．学会等名
情報処理学会音声言語情報処理研究会

４．発表年
2017年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	持橋　大地 (Daichi Mochihashi) (80418508)	統計数理研究所・数理・推論研究系・准教授 (62603)	