

令和元年6月13日現在

機関番号：34315

研究種目：若手研究(A)

研究期間：2015～2018

課題番号：15H05319

研究課題名(和文) ベイズ二重分節解析と深層学習の統合による教師なし言語獲得の構成論

研究課題名(英文) Unsupervised Language Acquisition Integrating Bayesian Double Articulation Analyzer and Deep Learning

研究代表者

谷口 忠大 (Taniguchi, Tadahiro)

立命館大学・情報理工学部・教授

研究者番号：80512251

交付決定額(研究期間全体)：(直接経費) 14,230,000円

研究成果の概要(和文)：本研究ではベイズ二重分節解析器の研究を発展させ、人間の実音声からの音素と語彙獲得を実現することを目的として研究を行った。主に三つの成果を挙げた。(1) 深層学習の活用によって実音声データからの教師なし語彙獲得の精度向上を実現した。(2) ベイズ二重分節解析器のサンプリングアルゴリズムの高速化により約百倍の高速化を果たし、二重分節解析に基づくロボットの自律的言語獲得へと道筋を立てた。(3) 深層学習を用いた教師なし話者適応による複数話者からの教師なし言語獲得手法を開発した。

研究成果の学術的意義や社会的意義

ロボットや人工知能が音声言語を認識するためには従来、大量の音声データとその書き起こし文を事前に人手で準備し、機械学習によって音声認識装置を訓練する必要があった。しかし、これは人間の幼児の言語獲得過程と一致しない。人間の幼児は書き言葉を学ぶ前に、音声データから音素や単語を発見していく。この過程を実現するために、本研究では研究代表者が構築してきたベイズ二重分節解析器に深層学習を組み合わせ、高い精度での音素・単語獲得、複数話者からの学習、また、そのアルゴリズムの高速化を実現した。

研究成果の概要(英文)：In this study, we extended the Bayesian double articulation analyzer aiming at achieving automatic phoneme and word discovery from real human speech. There were three main achievements. (1) We have realized the improvement of the accuracy of unsupervised phoneme and word acquisition from actual speech data by utilizing deep learning. (2) We developed an accelerated Bayesian double articulation analyzer that is 100 times faster than the previous algorithm. (3) We developed an unsupervised language acquisition method from multiple speakers by developing unsupervised speaker adaptation method using deep learning.

研究分野：創発システム

キーワード：記号創発 言語獲得 ノンパラメトリックベイズ 深層学習 音声認識

## 1. 研究開始当初の背景

音声言語には二重分節構造が含まれると言われる。二重分節構造とは古くから言語学において指摘されていた構造で、音声言語においては連続的な音声時系列情報が音素、単語という形で二重に分節化されることを指す。重要な点は音素(a, i, d 他)という有限個の単純なパターンの組み合わせにより、単語(dog, toy 他)という語彙に対応する無限個のパターンが生成されることである。単語は通常基本となる音素の確定的な並びにより作られ、また単語の並びは確率的にはあるが、規則的に並べられることで文を形成する。時系列データに潜む二重分節構造を推定することは、時系列データの中に隠れた潜在的な文字(音素)や潜在的な単語を推定することに等しい。音声言語の二重分節解析とは、単語とその確率的遷移ルールを獲得するという範囲で言語獲得を実現することを意味する。

研究代表者らは音声言語のみならず人間の動作系列や自動車運転行動のような時系列データにも二重分節構造が含まれるという仮説の下で、二重分節構造をベイズ教師なし学習により推定する二重分節解析器を提案し、一連の研究を行ってきた。二重分節解析器としては、ノンパラメトリックベイズ法に基づくHMM(隠れマルコフモデル)の無限状態変数拡張であるHDP-HMM(階層ディリクレ過程隠れマルコフモデル)とNPYLM(入れ子 Pitman-Yor 言語モデル)を用いたベイズ教師なし形態素解析器(チャンク化手法)を統合的に用いて構築した。この二重分節解析器を用い、人間の動作の分節化、自動車運転行動の分節化、ロボットの言語獲得に関する研究成果をあげてきた。特に自動車運転行動の分析については、意味ある運転行動単位への分割、運転動画要約、運転行動及び変化点の予測、運転トピックモデリングなどと幅広い研究成果をあげてきた。しかし、この二重分節解析器を音声言語に適用し、言語獲得を実現するには、HMMによる分節化部分とNPYLMによるチャンク化部分が確率モデルとして一体化されていないために、音素認識誤りに対応できないという大きな問題があった。この問題を解決するために、申請者らは「ノンパラメトリックベイズ理論に基づく生成的二重分節解析器」(科研費:若手BH24-H26)を通じて、二重分節構造の解析を単一の確率的生成モデルとして定式化したベイズ二重分節解析手法を構築した。研究代表者らの提案したベイズ二重分節解析器により言語獲得の制度は飛躍的に改善したものの実用的なレベルでの教師なし言語獲得は実現出来ていなかった。

## 2. 研究の目的

本研究では、これをさらに発展させて、ベイズ二重分節解析器の研究を進展させ、人間の実音声からの音素と語彙獲得を実現することを目的とする。この目的のためには(1)母音のみを対象としても精度が十分でなく、子音を含めた言語獲得が困難、(2)計算量が大きい(特に単語の最大時間長に対して3次のオーダー)、(3)音素間の違い以上に、話者間の音響特徴の違いを区別するように学習してしまう、といった問題を少なくとも解決する必要がある。

これらの問題を解決するために、ベイズ二重分節解析器自体の改善と、深層学習(ディープラーニング)の活用と統合による教師なし学習システムの構築を目指す。これにより、音素モデルも言語モデルの事前設計を一切必要とせず、人間の自然な音声発話データから発達かつ直接的な言語獲得を自律ロボットにおいて実現する機械学習手法を実現する。

研究期間内に、人間の大規模で連続的な音声発話時系列データから、その裏に潜む二重分節構造を推定し、音素モデルと言語モデルを獲得する学習器を構築する。その上で以下のように段階的に複雑性の高い音声言語時系列データに適用可能なように拡張し、ロボットによる教師なし言語獲得を実現することを目指した。

## 3. 研究の方法

本研究では、研究目的に示した教師なし言語獲得の計算論を構築するために、(1)ベイズ二重分節解析器と深層学習の統合モデルの開発、(2)ベイズ二重分節解析器のサンプリングアルゴリズムの高速化、(3)限られた語彙を用いた単一話者の自然な音声発話からの教師なし言語獲得の実現、(4)環境・文脈の情報を用いた教師なし言語獲得の高精度化、(5)複数話者・大規模音声発話データからの教師なし言語獲得の実現、を順次おこなう。二重分節構造を教師なし学習で分析する完全なベイズ生成モデルであるベイズ二重分節解析器と、教師なし学習による特徴抽出器である深層学習を統合させた上で、計算時間の問題を高速化アルゴリズムの開発により解決し、実音声データ解析とロボットの言語獲得への有用性を明らかにする。さらに、環境・文脈情報の活用により言語獲得の高精度化を行う。

## 4. 研究成果

以下に三項目に分けて主な研究成果を述べる。

### (1) 深層学習の活用による実音声データからの教師なし語彙獲得

本研究で対象とする教師なし言語獲得手法の開発のためには、ベイズ二重分節解析器(Nonparametric Bayesian Double Articulation Analyzer: NPB-DAA)の出力分布に適切な特徴ベクトルを受け取らせる必要がある。本研究では、深層学習器に音声時系列データから教師なし学習による表現学習を行わせ、適切な特徴ベクトルを抽出させるように学習させることを

目的として研究を行った。初年度は深層学習器の一種である Deep Sparse Auto Encoder (DSAE) を用いて教師なし学習により音声特徴量を抽出させることで、教師なし語彙学習の制度を向上させることを目指した。具体的には NPB-DAA と DSAE を結合させた NPB-DAA with DSAE を提案し、日本語話者に発生させた実母音系列からの単語獲得において既存手法を遥かに上回る性能を得た。母音のみによって構成された小規模の実音声データセットに関しては、既存の隠れマルコフモデルとラベルデータに基づく教師あり学習以上のパフォーマンスを示した[1]。

また、この研究を発展させ、子音まで含んだ実音声発話データに適用した。この際、子音と母音とでは音響特性が異なるために、どのように DSAE のネットワーク構造を用いるかが問題となる。これに対して比較研究を行い、子音を含んだ英語音声データに適用し、その有効性を検証した。具体的には TIDIGITS コーパスと呼ばれる数字の読み上げ音声データに適用し、その有効性を検証した[2]。

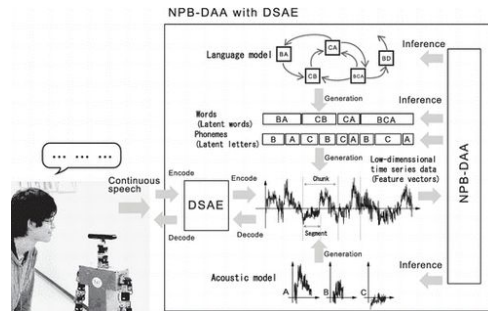


図 NPB-DAA with DSAE の概要図 [1]

## (2) ベイズ二重分節解析器のサンプリングアルゴリズムの高速化

従来のベイズ二重分節解析器のサンプリング速度は非常に遅く、人間の自然な音声発話を収集したデータからの解析では、非常に遅くスケールしない問題があった。これに対して動的計画法において用いられるメモ化の手法を NPB-DAA に適用する方法を考案し、効率的なメモリ利用によりオーダを一つ落とし、また、実装面での高速化により実質百倍のアルゴリズムの高速化を実現した[3]。

## (3) 深層学習を用いた教師なし話者適応による複数話者からの教師なし言語獲得の実現

(1) で導入した NPB-DAA with DSAE を複数話者のデータに適用すると、音素や単語の識別器が構築される以前に話者が識別されてしまうという問題があった。これはラベルデータを活用できる教師あり学習における問題以上に、教師なし学習による音素や語彙獲得を行う当該研究課題では解決の難しい問題である。これに対して、話者の識別情報が得られると仮定した上で、教師なし学習により話者非依存の特徴量を抽出する DSAE-PBHL (DSAE with Parametric Bias in Hidden Layer) を提案し、複数話者から話者非依存の特徴量を抽出し教師なし音素・語彙獲得が可能なることを示した(投稿中)。

## < 引用文献 >

[1] [Tadahiro Taniguchi](#), Ryo Nakashima, Hailong Liu and Shogo Nagasaka, Double Articulation Analyzer with Deep Sparse Autoencoder for Unsupervised Word Discovery from Speech Signals, *Advanced Robotics*, 30, (11-12), 770-783, 2016. DOI:10.1080/01691864.2016.1159981

[2] Yuki Tada, Yoshinobu Hagiwara, [Tadahiro Taniguchi](#), Comparative Study of Feature Extraction Methods for Direct Word Discovery with NPB-DAA from Natural Speech Signals, *IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob)*, 2017, DOI: 10.1109/DEVLRN.2017.8329802

[3] Ryo Ozaki, [Tadahiro Taniguchi](#), Accelerated Nonparametric Bayesian Double Articulation Analyzer for Unsupervised Word Discovery, *The 8th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics 2018*, 2018, pp. 238-244,

## 5. 主な発表論文等

### [雑誌論文](計2件)

[Tadahiro Taniguchi](#), Ryo Nakashima, Hailong Liu and Shogo Nagasaka, Double Articulation Analyzer with Deep Sparse Autoencoder for Unsupervised Word Discovery from Speech Signals, *Advanced Robotics*, 30, (11-12), 770-783, 2016. DOI:10.1080/01691864.2016.1159981

[Tadahiro Taniguchi](#), Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh, Symbol Emergence in Robotics: A Survey, *Advanced Robotics*, 30,

[学会発表](計 9 件)

Ryo Ozaki, Tadahiro Taniguchi, Accelerated Nonparametric Bayesian Double Articulation Analyzer for Unsupervised Word Discovery, The 8th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics 2018, 2018, pp. 238-244

Yuki Tada, Yoshinobu Hagiwara, Tadahiro Taniguchi, Comparative Study of Feature Extraction Methods for Direct Word Discovery with NPB-DAA from Natural Speech Signals, IEEE International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EpiRob), 2017, DOI: 10.1109/DEVLRN.2017.8329802

多田 裕貴, 幸 優佑, 林 楓, 萩原 良信, 谷口 忠大, ノンパラメトリックベース二重分節解析器の TIDIGITS コーパスへの適用, 2017 年度 人工知能学会全国大会 (第 31 回) (JSAI2017), 2017, 4D2-0S-37d-3

林楓, 岩田具治, 谷口忠大, 深層混合モデルによるクラスタリング 第 20 回情報論的学習理論ワークショップ (IBIS2017) ディスカッショントラック, 2017

Tadahiro Taniguchi, Kaede Hayashi, Ryo Nakashima and Shogo Nagasaka, Hierarchical Dirichlet Process-Hidden Language Model for Unsupervised Word Discovery from Speech Signals, Practical Bayesian Nonparametrics workshop at NIPS 2016, 2016

林楓, 中島諒, 長坂翔吾, 谷口忠大, 階層ディリクレ過程隠れ言語モデルと 深層学習を用いた語彙獲得過程の計算論, 日本認知科学会第 33 回大会, 2016

林楓, 高野敏明, 萩原良信, 谷口忠大, 階層ディリクレ過程隠れ言語モデルへの変分ベイズ法の適用, 2016 年度人工知能学会全国大会, 2016

中島諒, 谷口忠大, 長坂翔吾, 深層学習とノンパラメトリックベース二重分節解析器を組み合わせた実音声母音列からの教師なし語彙獲得, 計測自動制御学会 システム・情報部門学術講演会 2015 (SSI2015), 2015

中島諒, 谷口忠大, 長坂翔吾, ノンパラメトリックベース二重分節解析器と深層学習を組み合わせた母音列からの教師なし語彙獲得, 第 33 回 日本ロボット学会学術講演会, 2015

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。