

令和元年6月11日現在

機関番号：13901

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K16112

研究課題名（和文）行動文法の生成規則と発展様式の定量的理解

研究課題名（英文）Quantitative Study of Generative and Dynamical Principles in Action Grammars

研究代表者

笹原 和俊（Sasahara, Kazutoshi）

名古屋大学・情報学研究科・講師

研究者番号：60415172

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：本研究の目的は、行動系列の規則（行動文法）を分析するための新しいネットワーク手法を提案し、実データを用いて行動原理を定量的に理解することである。提案手法を鳥類のさえずりのデータおよびTwitterのデータを用いてその有効性を検証した。まず、オオムジツグミモドキを含む3種の鳥類のデータに提案手法を適用し、さえずりの系列規則を定量化するのに有効であることがわかった。さらに、Twitterの系列データを用いて、SNSユーザの属性推定やフェイクニュースの推定に応用したところ、どちらの場合も高い精度で推定できることがわかった。本研究によって、多様な行動データを分析するための基礎を確立することができた。

研究成果の学術的意義や社会的意義

人間や動物の行動は様々な行動要素が線形に組織化された系列データとして表現される。したがって、行動の複雑性や多様性を生み出すメカニズムを理解するためには、行動系列のもつ構造をいかに定量化するかが大きな鍵となる。しかし、既存の手法では適用に限界があった。そこで本研究では、行動文法の構造をモデル化するための新しいネットワーク手法を提案し、実データ（鳥類のさえずりとTwitterのデータ）を用いてその有効性を示した。DNAの構造理解が生命の理解に貢献したように、提案手法に基づく行動の構造理解は動物行動や人間行動の本質的理解に貢献することが期待できる。

研究成果の概要（英文）：The purpose of this study is to propose a new network method for analyzing the rules of behavioral sequences (action grammars) and to quantitatively understand the principles of behavior using real data. The proposed method was validated using bird song and Twitter data. First, we applied the proposed method to the data of three bird species including California Thrasher, and found that it was effective for quantifying the regularity of song sequences. In addition, we applied it to Twitter to infer the attributes of SNS users and detect fake news, and found that the accuracy of both estimates was high. This study has established a basis of the analysis of a wide variety of behavioral data.

研究分野：計算社会科学

キーワード：行動文法 時系列 ソーシャルデータ ネットワーク

## 様式 C-19、F-19-1、Z-19、CK-19 (共通)

### 1. 研究開始当初の背景

本研究の学術的背景には、計測や分析に関する情報技術の発展によって、行動の研究が新しい局面を迎えたことがあげられる。鳥類やクジラの歌、ハチのダンス、マウスの採餌行動、サルの道具使用、人間の言語コミュニケーション、これらはすべて多様な行動要素を組み合わせることで線形に組織化することで成立する系列行動である。最近の研究において、行動系列を定量的に理解することの重要性が論じられるようになった (Hauser et al., 2002)。特に、鳥類のさえずり (歌) に「文法」があることが発見されたのがきっかけとなり、行動の系列規則、すなわち「行動文法」の研究が注目を集めるようになった。

これまで研究代表者は、情報学のツールを用いて鳥類の歌の系列規則 (歌文法) を抽出する手法を提案してきた。例えば、オートマトン理論を用いて Bengalese Finch (ジュウシマツ) の歌文法を精度よくモデル化できることを実証し (Kakishita and Sasahara et al., 2009)、ネットワーク理論を用いて California Thrasher (オオムジツグミモドキ) の複雑な歌文法がスモールワールド構造をもつことを証明した (Sasahara et al., 2012)。一方、既存手法の限界もまた明らかになってきた。その顕著な例は時間変化する動的な系列データで、そのようなデータの分析には既存手法は有効ではない。

以上のような学術的背景から、本研究の着想が生まれた。

### 2. 研究の目的

人間や動物の行動は様々な行動要素が線形に組織化された系列データとして表現される。したがって、行動の複雑性や多様性を生み出すメカニズムを理解するためには、行動系列のもつ構造をいかにして定量化するのかが大きな鍵となる。しかし、既存の手法では適用に限界があった。

本研究の目的は、行動系列の規則 (行動文法) をモデル化するための新しいネットワーク手法を提案し、実データを用いて行動原理を探求することである。提案手法は行動パターンの生成機構と発展様式を探求するための強力なツールとなる。

### 3. 研究の方法

本研究では、行動系列データとして鳥類のさえずり (歌) とソーシャルメディア (Twitter) やネットの大規模言語データを使用し、提案手法を評価する。

- (1) 鳥類の歌は学習で獲得される複雑な行動系列のため、情報学的にも興味深い系列データを提供する。本研究では、鳥類の歌が音要素タイプをノード、その間の遷移をリンクとして系列データからネットワークに変換できることに着目して (Sasahara et al., 2012)、動的な系列データを取り扱うための処理を加えて、これまでのネットワーク手法を拡張する。そして、実データを用いて鳥類の歌文法を定量的に調査する。
- (2) ネットワーク手法の応用範囲は鳥類の歌にとどまらず、ハチのダンスやサルの道具使用、さらには言語行動まで多様な行動系列に適用可能である。そこで、代表的な SNS の 1 つである Twitter のタイムラインやウェブのニュースなどのデータを大規模に収集し、オンラインにおける言語コミュニケーションを提案手法で分析する。それによって、提案手法がオンラインの言語データから、どのような知識を発見することができるかを調査する。

### 4. 研究成果

以下、ネットワーク手法を 3 つ実データに適用した研究結果について報告する。

- (1) 分析に使用する行動系列データとして、2016 年夏に Charles Taylor 教授 (カリフォルニア大学ロサンゼルス校) と Ofer Tchernichovski 教授 (ニューヨーク市立大学ハンター校) の研究室を訪問し、California Thrasher (オオムジツグミモドキ)、Black-headed Grosbeak (チャバライカル)、Bengalese finch (ジュウシマツ) などの鳥類の歌データを収集した。そして、収集した歌データに提案手法を適用し、性能評価をおこなった。

その一例として、図 2 は、チャバライカルの歌の系列規則をネットワークでモデル化したものである。ノードは音要素の種類 (大きさは出現頻度に比例)、リンクは音要素間の遷移を示している。他の鳥類においても、さえずりの系列規則を定量化するのに有効であること

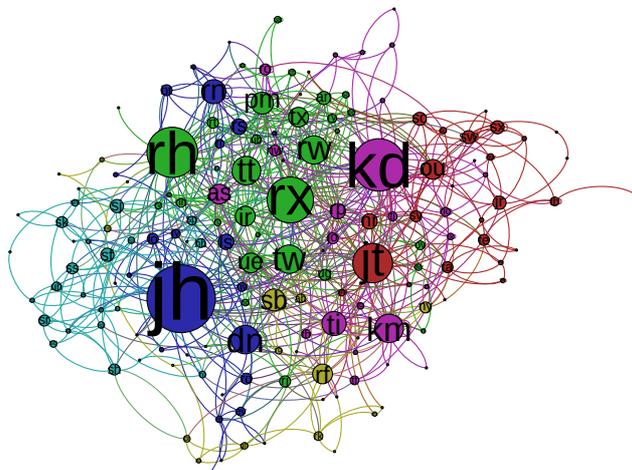


図 1 チャバライカルの歌文法  
(K. Sasahara (in prep.))

がわかった。ただし、音要素の種類が少なかったり、データサイズが小さかったりすると、提案手法は正確に行動系列を再現できないという問題も明らかになった。

- (2) Twitter のタイムラインのデータを用いて、ネットワーク手法によってどのぐらい個人属性を予測できるのかを調査した。個人属性の正解データは、著名人の場合は Wikipedia などのウェブの情報を確認し、一般のユーザの場合はアンケート調査を行い、その回答を正解とした。

その結果、性別（男か女か）、年齢層（デジタルネイティブか否か）、職業（10 種類）などの基本属性だけでなく、飲酒するか否か、Facebook 上の友人の多寡（ダンバー数 150 より多いか少ないか）、専門書を読むかどうかなど、多様な個人属性をネットワーク手法（階層的ニューラルネットワーク）によって、6 割以上の正答率で予測できることが明らかになった（図 2）。

興味深いことに、ツイートの時間順序をランダム化すると、正答率の平均自体は下がらないものの、分散が大きくなることがわかった。さらに、性別予測を例として、どのような情報を手掛かりとして、階層的ニューラルネットワークが識別を学習しているのかを調査したところ、男性も女性も使用する共通単語のうち性別と意味的に関連するものが重要な役割を果たすことがわかった（例：「バイク」の場合、男性は「KAWASAKI」や「一人旅」、女性は「ピンク」や「かわいい」が関連）。

- (3) 最後に、ネットワーク手法（階層的ニューラルネットワーク）が、言語的な手がかりを使って、偽ニュースの判別をすることが可能かどうかを実験した。本研究では、Snopes（偽ニュース全般を扱うサイト）と Politifact（政治系の偽ニュースを扱うサイト）で公開されているニュース記事と偽ニュース記事を収集し、機械学習のための正解データとした。

図 3 は、複数の機械学習による分類の結果を表したものである。Snopes のデータでは、サポートベクターマシン (SVC) の判別精度が約 80%、ランダムフォレスト (RF) とニューラルネットワーク (NN) の判別精度が約 90% だった。一方、Politifact の場合はロジスティック回帰 (LR) の精度が低くなっているものの、他の手法では識別精度が約 75% だった。これらの結果から、ネットワーク手法が常にベストとは限らないものの、安定して通常のニュースと偽ニュースを判別ができることが明らかになった。

以上の行動系列のデータ分析を通じて、ネットワーク手法の有効性といくつかの問題点が明らかになった。計測技術がますます発展し、行動のビッグデータの新しい活用方法の発見が求められる今、本研究の成果はそのための重要な基礎となる。

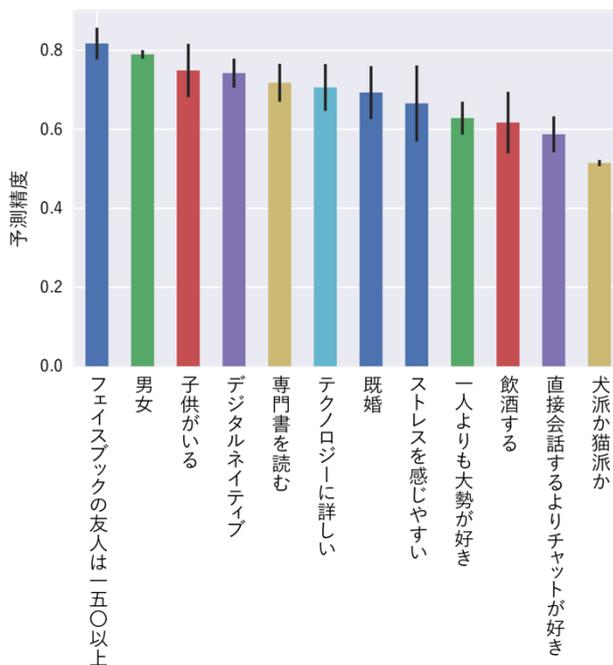


図 2 : Twitter の投稿からの個人属性予測 (笹原 2018)

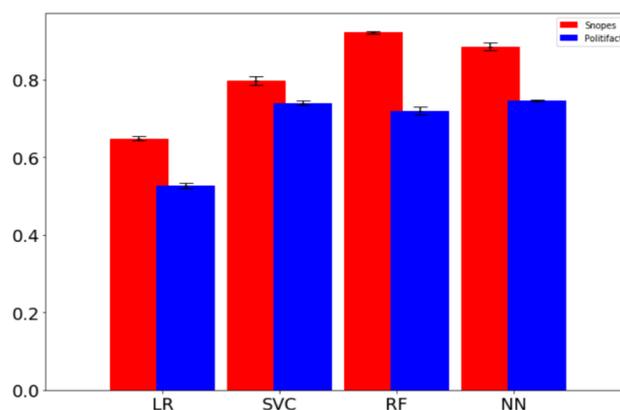


図 3 : 機械学習による偽ニュースの判別  
LR : ロジスティック回帰, SVC : サポートベクターマシン,  
RF : ランダムフォレスト, NN : ニューラルネットワーク,  
赤 : スノープス, 青 : ポリティアクト

K. Sasahara (in prep.)

## 5. 主な発表論文等

### 〔雑誌論文〕(計 4 件)

1. 笹原和俊, ビッグデータとは何か, 社会と調査 22, 査読無, pp. 8-15, 2019
2. K. Sasahara, You are what you eat: A social media study of food identity, Journal of Computational Social Science, 査読有, pp.1-15, 2019, <https://doi.org/10.1007/s42001-019-00039-7>
3. 笹原和俊, つぶやきの中のモラル, 人工知能 34(2), 査読無, pp. 146-151, 2019
4. T. Yo and K. Sasahara, Inference of Personal Attributes from Tweets Using Machine Learning, Proceedings of the 2017 IEEE Big Data, 査読有, pp. 3086-3092, 2017, 10.1109 / BigData.2017.8258295

### 〔学会発表〕(計 9 件)

1. 笹原和俊, フェイクニュース現象の計算社会科学, 第3回計算社会科学ワークショップ, 2019.03.02
2. T. Yo and K. Sasahara, Prevalence of Personal Attributes Found in Twitter Posts, 第3回計算社会科学ワークショップ, 2019.3.3
3. 笹原和俊, フェイクニュース現象の計算社会科学, 第3回計算社会科学ワークショップ, 2019.03.02
4. 余岳, 笹原和俊, ツイートが運ぶ多様な個人属性の定量化, WiNF2018, 2018.11.10
5. 杉森真樹, 笹原和俊, 虚偽情報の拡散現象の定量化: 言語的特徴と拡散基盤の分析, WiNF2018, 2018.11.10
6. M. Sugimori, K. Sasahara, Analyzing behavioral complexity of social bots and bot clusters, EVOSLACE in ALIFE2018, 2018.7.25
7. T. Yo, A. Saji, K. Sasahara, Quantification of Diverse Personal Attributes in Tweets, 人工知能学会第32回全国大会, 2018.6.6
8. 杉森真樹, 笹原和俊, 時田恵一郎, ソーシャルボットの検出: 言語非依存性の特徴量とボット集団の定量化著書, 人工知能学会第32回全国大会, 2018.6.5
9. 余岳, 笹原和俊, 機械学習と深層学習を用いたソーシャルデータからの個人属性の推定, 第11回テキストアナリティクス・シンポジウム, 2017.9.8

### 〔図書〕(計 1 件)

1. フェイクニュースを科学する 拡散するデマ、陰謀論、プロパガンダのしくみ, 化学同人, 192, 2018

### 〔その他〕

ホームページ

<https://www.colorlessgreen.info>

メディア

1. 沖縄タイムスに「エコーチェンバー現象」の仕組みについて、研究成果を交えながら解説した。(2019.3.21, 22)
2. FIJ セミナー「フェイクニュース現象の本質は何か」(2019.3.10)に関する記事が掲載された。
  - ・ 琉球新報(2019.03.11)
  - ・ BuzzFeed JAPAN (2019.03.10)
3. NHK 総合テレビの番組「クローズアップ現代+」にインタビュー形式で出演した。(2019.3.4)
4. 『フェイクニュースを科学する 拡散するデマ、陰謀論、プロパガンダのしくみ』の書評が掲載された。
  - ・ 京都大学生協 書評雑誌「綴葉」(2019.3.25)
  - ・ 日刊工業新聞 (2019.3.18)
  - ・ 公明新聞 (2019.2.18)
  - ・ 週刊新潮 2019年2/14号 (2019.2.8)
  - ・ 京都新聞 (2019.1.5)
  - ・ 日本経済新聞 (2018.12.15)
5. 琉球新報「ボット」の調査結果に関して、欧米におけるボット研究の知見を踏まえて現状と今後の課題、可能な対策について掲載された。(2019.1.1, 2018.12.31)

6. 科学雑誌ニュートン「SNSのエコーチェンバー現象を再現するコンピュータシミュレーション」が紹介された。(2018年7月号)

#### アウトリーチ

1. シンポジウム「フェイクニュースとどう向き合うかーメディアの現場と社会科学・情報科学の対話」で講演した。(2019.3.31)
2. JST CREST「脳領域／個体／集団間のインタラクション創発原理の解明と適用」第2回公開シンポジウムで講演した。(2019.3.16)
3. FIJセミナー「フェイクニュース現象の本質は何か」で講演した。(2019.3.10)
4. 価値創造研究センターシンポジウム「ポジティブ情報学の誕生」で講演した。(2019.3.4)
5. マル激トーク・オン・ダイヤモンドに出演し、フェイクニュース問題に関して討論した。(2018.12.15)
6. 名古屋大学情報学部・情報学研究科 公開セミナーで講演した。(2018.11.17)
7. 三省堂書店名古屋本店 第8回学問喫茶で講演した。(2018.11.13)
8. 平成30年度名古屋大学公開講座で講演した。(2018.10.16)
9. 群馬大学第4回計算社会科学とその周辺セミナーで講演した。(2018.7.19)
10. 夢ナビライブ2018で講演した。(2018.6.16)

## 6. 研究組織

### (1) 研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

### (2) 研究協力者

研究協力者氏名：Charles Taylor

ローマ字氏名：チャールズ・テイラー

カリフォルニア大学ロサンゼルス校・教授

研究協力者氏名：Ofer Tchernichovski

ローマ字氏名：オファー・チェルニコフスキー

ニューヨーク市立大学ハンター校

研究協力者氏名：Filippo Menczer

ローマ字氏名：フィリッポ・メンツァー

インディアナ大学ブルーミントン校・教授

※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。