

## 科学研究費助成事業 研究成果報告書

平成 29 年 5 月 26 日現在

機関番号：12601

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26540133

研究課題名（和文）身体運動と環境の大規模実世界データの言語表現化

研究課題名（英文）Automatic description of human actions in the real world

研究代表者

高野 渉（Takano, Wataru）

東京大学・大学院情報理工学系研究科・准教授

研究者番号：30512090

交付決定額（研究期間全体）：（直接経費） 2,800,000円

研究成果の概要（和文）：実世界の連続情報から記号を抜き出して理解する人工知能が求められる。本研究課題では、人間の行動を言語として理解する計算論を開発した。身体運動や操作物体を記号化し、その記号と言語を結び付ける機械学習によって行動から文章を生成する技術である。特に、運動の時系列や文章の遷移から文脈を抽出することによって、行動の背後にある文脈を推定しながらそれに即した文章を生成することを可能にした。

研究成果の概要（英文）：Artificial intelligence is expected to symbolically understand the real world. In this research, we have focused the human actions, and developed the fundamental framework to convert the actions to their relevant descriptions. This framework encodes human whole body motions and objects to be manipulated, and subsequently establish the connection between these symbols and the descriptions. This connection allows for generating sentences describing the human actions. Especially, the extraction of the contexts from sequences of the actions and sentences leads to selecting words appropriate to the current situations, and results in making correct descriptions for the human actions.

研究分野：ロボティクス

キーワード：行動認識 機械学習 統計モデリング

1. 研究開始当初の背景

実世界は膨大な連続情報で満たされている。人間は、その連続情報の中から情報を取捨選択し、それを記号に落とし込んで認識している。記号を道具として豊かな連想を膨らませることが人間の高次な知能の根源にある。これまで、人間の身体運動を統計モデルとして記号化することによって、人間の行動を記号として識別する研究は行われてきた。さらに、その記号と自然言語が結び付くことによって、人間の身体運動が記号を介して文章として認識される人工知能技術へと発展してきている。しかし、必ずしも人間の行動は時々刻々の身体運動のみから言語化できるものではない。人間は、身体運動、環境、言語表現を有機的に結びつけた膨大な経験と知識を活用しながら行動を理解している。行動と言語のビッグデータを構造化し、それを再利用することによって、実世界での人間の活動を言語として理解する人工知能が求められている。

2. 研究の目的

研究代表者はこれまで、人間の全身運動を計測し、それを隠れマルコフモデル(HMM)として学習する知能の枠組みを開発してきた。これは、人間の運動を HMM として識別することと、HMM から人間と同じような全身運動を生成することを機能としたヒューマノイドロボットの人工知能である。さらに、運動を学習した HMM を運動の連続データを圧縮した記号としてみることによって、この運動の記号と言語を結ぶ統計数理モデルを開発してきている。しかし、運動単位ごとにしか記号化することができず、運動以外のモーダル情報を取り込めていないため、行動の言語表現は画一的で無味乾燥にならざるを得なかった。身体運動とそれを取り巻く環境、さらにそれらの時間的変遷が行動の意図や目的に影響を及ぼす。本研究では、これまでの運動と言語を結び付ける統計数理モデルを、操作物体や時間変遷を表現できるように拡張することによって、行動の文脈を推定しながら、その文脈に応じて行動を言語として表現する計算論を構築することを目的とする。

3. 研究の方法

行動と言語の大規模なデータから文脈を統計モデルとして表現し、文脈を推定しながら行動を言語化するため枠組みを構築する。以下の研究開発を実施した。

(1) 大規模言語コーパスから行動の文脈を表現する技術

身体運動を HMM として学習した運動記号とその運動に付けられた文章中の単語を結び付ける統計モデルを開発してきている。これは運動の記号から隠れ状態が生成され、隠れ状態から単語が生成される統計モデル

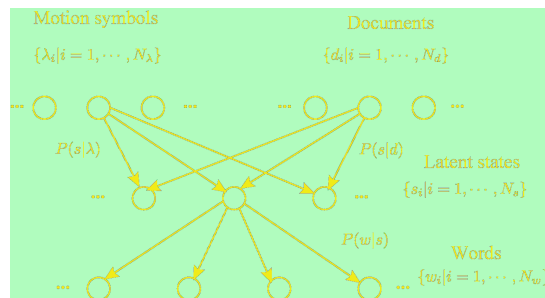


図1. 大規模言語コーパスからのトピック抽出とトピックを介した運動と単語の統計モデル

である。このモデルを図1のように、文書(文章の集まり)から隠れ状態が生成され、隠れ状態から文書に含まれている単語が生成される構造を埋め込む。運動と単語の関係と文書と単語の関係を隠れ状態を共有した統計モデルとして学習することによって、隠れ状態が文書データに埋め込まれた文脈を表現することになる。この統計モデルと別に運動に付与された文章や文書データセットから文章構造を単語の遷移確率として学習しておく。運動と単語の連想関係と、単語の繋がりを知識として活用することによって、運動から関係の深い単語を生成し、それら単語を文章中での繋がりを考慮して並べることによって、運動から文章を生成することが可能となる。

(2) 環境中の複数物体認識と運動認識との統合

身体運動だけでなく、身体が働きかける環境中での物体の識別を行動の言語化へ活用する枠組みを開発する。図1にあるような運動と単語の統計モデルの第1層(運動記号と同じレベル)に物体記号を取り込むようにモデルを拡張する。運動の記号とその複数の物体の記号との組み合わせに応じて、単語が生成される確率が求められる。行動データに応じて生成された単語の集合を、上述と同じ単語の遷移知識を利用して適切に並べることによって文章を作成することができる。

(3) 行動と文章の時系列の統計モデリング

一連の行動における運動記号の移り変わりや言語表現の変遷を表現する統計モデルを構築する。図1の統計モデルでは、隠れ状態間の遷移を無視している。隠れ状態(トピック)が遷移し、そのトピックに応じて運動や言語表現が生み出されるという仮定のもと、隠れ状態間の遷移確率を導入する。学習では、隠れ状態の遷移確率を、運動記号と文章の時系列を学習データとし、運動記号列から文章が生成される確率が最大となるように最適化する。この最適化はEMアルゴリズムを採用する。行動の文脈を隠れ状態(トピックのダイナミクス)として抽出・同定することがこの統計モデルの鍵である。

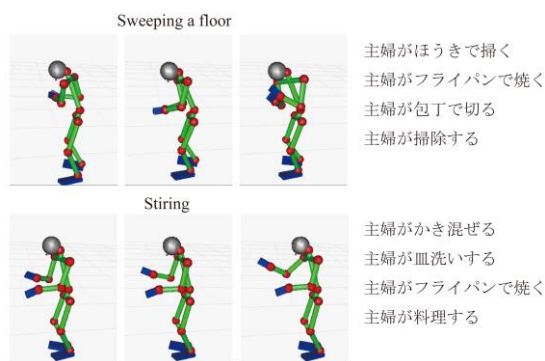


図 2. 全身運動から文章の生成実験結果

#### (4) 言語ダイナミクスの幾何学モデル

単語と運動の統計的な結びつきから単語間の(非)類似度を計算できる。類似度が大きい単語同士は近くに、類似度が小さい単語同士は遠くなるように単語を多次元空間上の点に配置した言語空間を構築してきた。この空間は、実世界の運動に根差した単語の関係性を幾何学的に捉えているが、単語のダイナミクス(単語の遷移)を表現していない。文章中での単語の遷移が、空間中で単語が等速かつ等曲率で移動する質点運動として表現できるように単語を空間中に配置するアルゴリズムを開発した。単語の運動はこれまでの行動や言語表現に依存し、時系列文脈を内在している。また、単語の運動として文章を捉えることによって簡単な幾何学計算によって単語を生成し、文章を作成することが可能となる。

#### 4. 研究成果

(1) 光学式モーションキャプチャにて計測した全身運動 470 個、運動に付けられた文章 764 個、および青空文庫の本 10,925 冊(段落を文書として扱い、291,191 文書)を学習した。学習データに含まれる語彙数は 111,998 であり、単語間の遷移として 23,382,822 個の単語 trigram を用いた。学習した統計モデルを用いて全身運動から文章を生成した実験の結果の一部を図 2 に示す。「ほうきで掃く」動作から、「主婦がほうきで掃く」「主婦が掃除する」という正しい文章が生成されている。言語コーパス中において、掃除に係わる文章が料理の文章と同じ文書で使用されることが頻繁にあるため、掃除と料理が同じ隠れ状態を共有し、料理に係わる「主婦がフライパンで焼く」、「主婦が包丁で切る」という文章も生成された。「かき混ぜる」動作からは、「主婦がかき混ぜる」という正しい文章と料理に関係する「主婦が皿洗ひする」、「主婦がフライパンで焼く」、「主婦が料理する」が生成されている。言語コーパスの知識を活用しながら、学習データ中で運動に付けられていない単語を用いた文章が生成できることを確認した。動作データから文章を 5 つ生成し、最も確率が高い文章と動作データに付けられた正解文章との一致率は 36.3%、5 つの文章のうちの 1 つが正解文章と一致

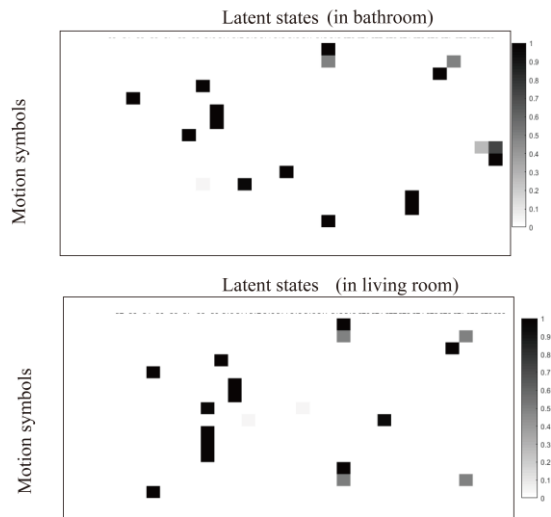


図 3. 運動データに応じた隠れ状態の生成分布

する割合が 78.3%であった。

(2) 身体運動とその周囲環境を RGBD カメラにて計測した。身体運動は、骨格を復元しその関節角度の時系列として表現される。物体を含む領域を切り出し、その領域の色分布からフィッシャベクトルを計算する。これを物体の特徴量とし、特徴量をクラスタリングすることで、物体データを記号化する。複数物体(11 種類)を操作する運動(22 種類)の行動を 396 回計測した。264 個の行動データを学習し、132 個のデータで行動データから文章の生成試験を行った。生成された文章と正解文章の単語 ngram の一致率を数値化した BLEU スコアは 0.41 であった。なお、従来の操作物体を 1 つに限定したモデルでは BLEU スコアは 0.58 であった。複数物体を考慮すると、運動と複数の物体から単語が生成される確率のパラメータ数が増え、それに応じて学習データを増やしていく必要がある。

(3) 約 30 秒の動作データを分節化し、分節ごとを運動記号に変換する。同じ分節に文章を付与する。約 20 時間程度の動作データに対して、このような運動記号列と文章列を学習した。洗面所と居間で計測した動作データを入力したときに、各運動記号から隠れ状態が生成される確率分布を図 3 に示す。動作データによって運動記号から生成される隠れ状態の確率が変わることを表しており、動作の状況に応じた文脈が隠れ状態に埋め込まれていることが推測される。このような文脈を隠れ状態の遷移として表現するモデルから洗面所での運動データから文章を生成した。図 4 に運動データから、「歩く」「立ち止まる」「ドアを開ける」「棚を開ける」「うがいする」「水を吐く」「水を捨てる」「手を洗う」「手を乾かす」を説明する文章が生成されることが確認できる。状況に応じた隠れ状態を推定し、そこから適切な単語が連想さ

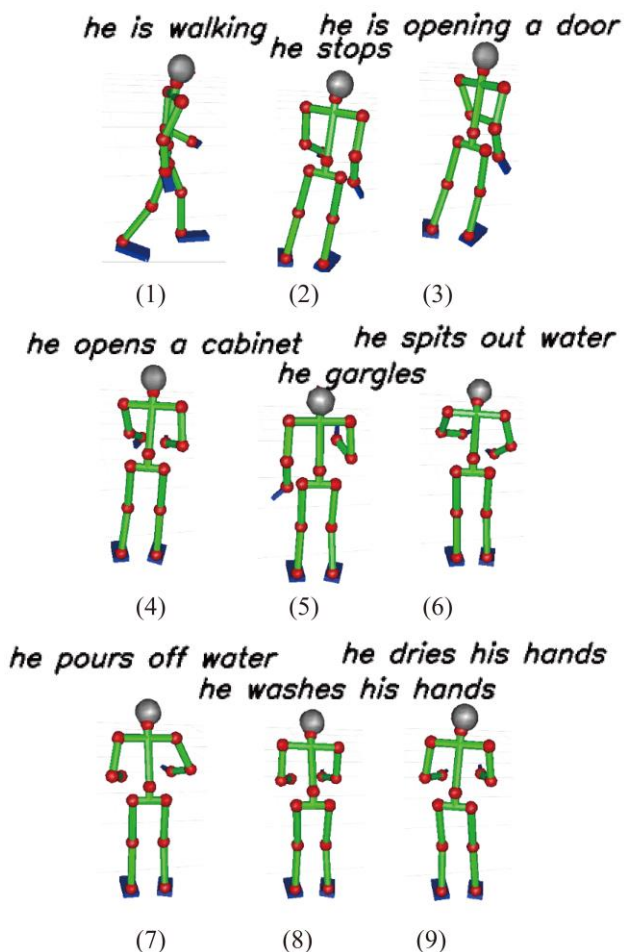


図4. 一連の運動からの文章生成

れていることを示している。

(4) 青空文庫の本8冊からその中に含まれる単語を多次元空間の点として布置した。文章中の単語の出現は空間上での軌跡として現れ、その軌跡が等間隔かつ局率が急激に変化しないように単語を配置した。単語の種類数は8,495個であり、空間の次元数は16と設定した。単語の軌跡に沿って次の単語を予測することができる。予測は、軌跡の局率を維持しながら、一定距離延長し、その終点近傍の単語を見つける計算になる。終点との距離が近い単語を5つ検索した。最も距離が近い単語と正解単語の一致率は、24.8%であり、検索した5つの単語に正解単語が含まれている割合は48.4%であった。また、この空間を利用して文章生成の実験を行った。空間を使って生成される文章とリカレントニューラルネットワークから生成される文章の定量評価を行った。生成した文章をクラウドソーシングに掲載し、ある文書に続くにふさわしい文章を選んでもらい、その回答を集計する。2000個の回答から空間を利用して生成された文章のほうが自然と回答されたものが71%であった。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 7件)

[1]王 佳辰, 高野 渉, 中村 仁彦, “行動理解のための言語コーパスからのトピック抽出”, 第34回日本ロボット学会学術講演会, 3W3-08, 山形大学, 山形県, 2016.9.7-9

[2]前川 知行, 高野 渉, 中村 仁彦, “文脈を空間的に表現する言語モデルを用いた文章生成”, 第21回ロボティクスシンポジウム, 5D3, 長崎やすらぎ伊王島, 長崎 2016.3.17-18

[3]前川 知行, 高野 渉, 中村 仁彦, “CECESS モデルによる文脈の可視化と予測”, 第33回日本ロボット学会学術講演会, 2B2-08, 東京電機大学, 東京都, 2015.9.3-5

[4]Yoshihiko Yamada, Wataru Takano, Yoshihiko Nakamura, “Statistical Behavioral Understanding by Motion, Object and Language”, The 14th IFToMM World Congress, OS13-099, Taipei, Taiwan, Oct 25-30, 2015

[5]Yusuke Goutsu, Wataru Takano, Yoshihiko Nakamura, “Gesture Recognition Using Hybrid Generative-Discriminative Approach with Fisher Vector,” IEEE International Conference on Robotics and Automation, pp.3024-3031, Seattle, USA, May 26-30, 2015

[6]山田 嘉彦, 高野 渉, 中村 仁彦, “身体運動と3次元物体の認識から行動を文章として理解するヒューマノイドロボットの知能”, 第32回日本ロボット学会学術講演会, 1I3-03, 九州産業大学, 福岡県, 2014.9.4-6

[7]山田 嘉彦, 高野 渉, 中村 仁彦, “身体と環境の幾何学的関係を抽出する物体操作記述子による運動の記号化”, 日本機械学会ロボティクスメカトロニクス講演会 2014, 3P2-Q06, 2P1-K05, 富山国際会議場, 富山市, 2014.5.25-29

## 6. 研究組織

(1) 研究代表者

高野 渉 (TAKANO WATARU)

東京大学・大学院情報理工学系研究科

・准教授

研究者番号：30512090

(2) 研究分担者

なし

(3) 連携研究者

なし

(4) 研究協力者

なし