身体運動と言語を結ぶロボットの統計的情報処理

高野 渉* 鮎澤 光* 濱野 聖也*
梅澤 慶介* 中村 仁彦*

Statistical Processing that Combines Whole Body Motion and Natural Language for Robots

Wataru Takano*, Ko Ayusawa*, Seiya Hamano*,
Keisuke Umezawa* and Yoshihiko Nakamura*

This paper describes a novel approach to linguistic mutual inference, which enables robots not only to linguistically interpret the motion patterns as sentences but also to recall the motions from the sentences. The inference can be established based on the motion language model and the natural language model. The motion language model stochastically connects the symbols of motion patterns to the words through the latent states which represent the underlying linguistic structure. The natural language model represents sequences of words. The motion language model and the natural language model correspond to semantics and syntax respectively. The integration of the motion language model and the natural language model allows the linguistic mutual inference for the robots. The two kinds of inference can be made by solving search problems: search for a sequence of words and a symbol of motion pattern. The proposed approach to interpretation of motion patterns as sentences and recall of motion patterns from the sentences through integration of the motion language model and the natural language model is validated by the experiment on the human behavioral data.

Key Words: Symbolization of Motion Patterns, Natural Language, Hidden Markov Models, Statistical Inference

1. はじめに

人間は進化の過程で言語を獲得してきた。実世界の事象は言語を通して切り分けられ、その記号の集合として理解される。これは、構造主義にも通じる思想であり、人間の認知は社会が共有する言語が作り上げた記号体系の上に成り立っているのである。言語・記号情報処理は、人間の高次な知能計算の根幹を成す。

Saussure は、言語の仕組みを「シーニュ」、「シニフィアン」、「シニフィエ」を用いて説明している[1]。記号（シーニュ）は指示するものである記号表現（シニフィアン）と指示されるものの概念を表す記号内容（シニフィエ）の対として構成される。その対の組み合わせは恣意的であり、この恣意性によって記号は実環境そのものの事象から切り離され、記号の操作性が向上する。その結果、記号を理解する仕組みである言語（ラング）が獲得されたとしている。

Deacon も実環境の事象と記号との結合関係に着目し、「アイコン」、「インデックス」、「シンボル」という記号の階層構造を通じた人間の知能の発達過程を提唱している[2]。物理事象との類似性に基づく参照記号である「アイコン」から物理事象との因果関係が薄れて記号間の結びつきが強まることによって「シンボル」へと記号が発達する。そしてこのような記号の発達過程を経て言語も生まれたと考えられている。

脳科学では、他者の特定の運動を観察したとき、および自身が同じ運動を行ったときに発火するミラーニューロンが発見され注目を集めている[3]。ミラーシステムと、行動の認識・生成、こころの理論、コミュニケーション、言語情報処理との関係性が指摘されている[4]。このように、運動の認識と生成を結び付けるミラーシステムは多様な役割を担っていると推測されている。

ロボティクスにおいても、身体感覚・運動の記号化に基づくロボットの知能構築に関する研究が行われている[5]~[10]。これらのロボットの記号情報処理のアプローチは認知心理学や脳科学で唱えられている記号論・言語論のパラダイムに接近しつつある[11]~[14]。Sugita らは運動と文章のニューラルネットを結合することによって文章から運動を生成する手法を提案した[11]。Ogata らは、この枠組みを発展させ、運動から文章を生成する計算法を開発した[12]。文章生成の計算は、運動から複数の文

原稿受付 2011 年 12 月 9 日

* 東京大学大学院情報理工学系研究科

* Graduate School of Information Science and Technology, University of Tokyo

■ 本論文は学術性で評価されました。

章を連想する過程と連想された文章から運動を生成し、元の運動と比較することによって適切な文章を選定する過程から構成される。そのため、文章生成には大きな計算負荷がかかるという問題がある。さらに、このニューラルネットの枠組みは運動と文章を結合する条件として新たな制約が加わるため、多数の運動と文章を学習することは困難である。コンピュータグラフィックスでは、運動と言語という観点から多様な運動を合成する研究が行われてきた。Arikan らは、言語から CG キャラクタの運動を合成する手法を提案している。データベース中の運動の各フレームに動詞のラベルが付与されている。入力された動詞に対応するフレームの運動を拘束条件として、それを満たしながら連続な運動を動的計画法を通じて求めることによって、言語に対応した新しい運動を合成することが可能である [15]。Rose らは類似した “Verb” と呼ばれる運動グループ内で二つの運動の差を表現するパラメータを “Adverb” として導入し、そのパラメータをコントロールすることで二つの運動を補間する新しい技術を提案している [16]。この “Verb” や “Adverb” という概念は導入されているが、実際の言語との関係性については触れられていない。

自然言語の分野にて、大規模な言語コーパスを扱う手法として、統計論的アプローチが広く用いられている。このアナロジーを利用して、実世界を記号として分類する計算からロボットの言語獲得を目的とする研究がなされている。Nakamura らは視覚・聴覚・触覚の身体を通じて取り込まれる物体に関する感覚情報を統計モデルによって分類する手法を提案した [10]。この分類構造と単語を結び付けることによって、単語の概念を形成する計算法へと発展している [14]。筆者らも運動パターンの記号と動詞の各時系列間の連想関係を統計モデルにて表現することにより、観察した運動から動詞への翻訳、および動詞から運動パターンの連想が可能であることを確認してきた [17] [18]。しかし、これらの枠組みでは、単語が実世界からの情報に接地し

ているが、文章の構造までも獲得するには至っていない。実世界情報としての運動データを記号化する情報処理および運動と多様な単語や文章構造を表現する自然言語情報処理を結びつける枠組みが必要である。前者は運動データをシニフィエ、運動の記号をシニフィアンと見た運動記号の設計手法であり、後者はその運動記号と言語の記号を階層的記号システムとして構築する手法である。

本稿では、運動パターンを表す運動記号から単語を連想する運動言語モデルと文章の構造を単語の並びとして表現する自然言語モデルを統合することによって、運動パターンから文章生成および文章から運動の連想を可能とするロボットの言語情報処理モデルを提案する。運動言語モデルは、非観測な情報を隠れ状態として表現し、運動パターン記号、隠れ状態、単語の階層的連想構造を統計的に表現したモデルとして与えられている。また、自然言語モデルとして、文章中の単語はその直前の単語にのみ依存するという仮定に基づいた単語 bigram モデルを採用した。運動言語モデルは、隠れ状態を介して単語と運動記号の関係性を表現する意味モデルに相当し、自然言語モデルは、単語の並びを表現する文法モデルに相当する。運動から文章の生成、文章から運動の連想を、運動言語モデルと自然言語モデルを用いた探索問題として解くことによって高速な言語処理が可能であることを示す。

2. 運動言語と自然言語のモデル化

運動を基盤としたロボットの言語処理モデルを構築するために、運動と単語の連想関係を統計的に表した運動言語モデルと単語の並びを表した自然言語モデルを提案する。Fig. 1 に示すように、運動言語と自然言語モデルを統合することによって、運動を言語として解釈することおよび言語から運動を連想することが可能となる。

Fig. 1 Overview of integration of a motion language model with a natural language model. The motion language model represents relationship among motion symbols and words via latent variables as a graph structure, where nodes on 1st, 2nd and 3rd layer indicate the motion symbols, the words and the latent variables respectively. The natural language model represents sentence structure as transitions between two words. The motion language model and the natural language model are equivalent to semantics and syntax processes. By integrating two processes, linguistic processing for robots can be realized

Fig. 2 The motion language model represent the stochastic association of words with motion symbols via latent states. The motion language is defined by two kinds of parameters : probability that a word is generated by a latent variable and probability that a latent variable is generated by a motion symbol

2.1 運動言語モデル

言語は音声や文字などの観測可能な情報と単語の意味や品詞などの非観測な情報を含んでいる。言語の非観測な情報を表現する手法として、隠れ状態を含む統計モデルは有効である。本稿では、運動パターンの記号と単語との連想関係を隠れ状態を介した統計モデルによって表現する。Fig. 2はこのモデルの概略図を表す。以下この統計モデルを運動言語モデルと呼ぶことにする。

ロボットは身体運動の時系列データを隠れマルコフモデル(HMM: Hidden Markov Model)によって記号化する。ロボットはHMMのパラメータ集合で表現される運動記号を記憶することによって運動知識を蓄えていく。記憶した運動記号を用いた統計的情報処理により観察した運動を認識すること、自身の運動を生成することが可能となる[9]。

ロボットはある運動を観察したとき、その運動を運動記号として認識する。人間により同じ運動を表す文章がロボットに教示される。運動記号としての認識と文章の教示を通じて、運動言語モデルを用いて運動記号と単語の連想構造が学習される。運動言語モデルは運動記号 λ から隠れ状態 s を連想する確率 $P(s|\lambda)$ と隠れ状態 s から単語 ω を連想する確率 $P(\omega|s)$ のパラメータから構成される。運動記号、単語、隠れ状態の集合を、おのこの $\{\lambda_i | i = 1, 2, \dots, N_\lambda\}$, $\{\omega_i | i = 1, 2, \dots, N_\omega\}$, $\{s_i | i = 1, 2, \dots, N_s\}$ と表現する。

ロボットが k 番めに観察した運動データの認識結果を運動記号 λ^k とする。同じ運動データに人手により付けられた文章(単語列)を $(\omega_1^k \omega_2^k \dots \omega_{n_k}^k)$ とする。ここで、 n_k は文章の長さを表している。この運動記号と文章の対の集合から以下の評価関数を設定する。

$$\Phi = \sum_{k=1}^N \log P(\omega_1^k, \omega_2^k, \dots, \omega_{n_k}^k | \lambda^k) \quad (1)$$

これは、観察した運動の認識結果である運動記号 λ^k から文章を構成する単語の集合 $\{\omega_1^k, \omega_2^k, \dots, \omega_{n_k}^k\}$ が生成される確率の和を表している。ここで、N は学習過程において与えられた運動記号と文章の対の総数である。

運動記号から単語が生成される確率はその運動記号にのみ依

存するという仮定を用いることによって、式(1)右辺の条件付確率を以下のように近似する。

$$P(\omega_1^k, \omega_2^k, \dots, \omega_{n_k}^k | \lambda^k) = \prod_{i=1}^{n_k} P(\omega_i^k | \lambda^k) \quad (2)$$

さらに、式(2)右辺の条件付確率は、運動言語モデルのパラメータである $P(s|\lambda)$ と $P(\omega|s)$ を用いて式(3)のように表すことができる。

$$P(\omega_i^k | \lambda^k) = \sum_{j=1}^{N_s} P(\omega_i^k | s_j) P(s_j | \lambda^k) \quad (3)$$

式(1)の評価関数が最大になるように、EM アルゴリズムによって運動言語モデルのパラメータを最適化する。このEM アルゴリズムは、前回推定されたモデルパラメータに基づき隠れ変数の分布を計算する E-Step と計算した隠れ変数の分布に基づき評価関数が最大となるようにモデルパラメータの最適値を推定する M-Step を交互に計算する手法である。EM アルゴリズムの詳細は付録 A に譲ることにする。

E-Step によって隠れ変数の分布は式(4)として求める。

$$P(s | \lambda^k, \omega_i^k) = \frac{P(\omega_i^k | s, \lambda^k, \theta) P(s | \lambda^k, \theta)}{\sum_{j=1}^{N_s} P(\omega_i^k | s_j, \lambda^k, \theta) P(s_j | \lambda^k, \theta)} \quad (4)$$

ここで、 θ は前回の M-Step にて推定されたモデルパラメータの集合である。式(4)において、分子は運動言語モデルを通じて運動記号 λ^k から単語 ω_i^k と隠れ変数 s が連想される同時確率、分母は運動記号 λ^k から単語 ω_i^k が連想される確率である。

M-Step によってモデルパラメータは以下のように推定される。

$$P(s | \lambda) = \frac{\sum_{k=1}^N \sum_{i=1}^{n_k} \delta(\lambda, \lambda^k) P(s | \lambda^k, \omega_i^k)}{\sum_{j=1}^{N_s} \sum_{k=1}^N \sum_{i=1}^{n_k} \delta(\lambda, \lambda^k) P(s_j | \lambda^k, \omega_i^k)} \quad (5)$$

$$P(\omega | s) = \frac{\sum_{k=1}^N \sum_{i=1}^{n_k} \delta(\omega, \omega_i^k) P(s | \lambda^k, \omega_i^k)}{\sum_{j=1}^{N_\omega} \sum_{k=1}^N \sum_{i=1}^{n_k} \delta(\omega_j, \omega_i^k) P(s | \lambda^k, \omega_i^k)} \quad (6)$$

ここで、 $\delta(\lambda_i, \lambda_j)$ および $\delta(\omega_i, \omega_j)$ はクロネッカーデルタを表し、 i と j が等しいとき 1 の値を取り、その他の場合は 0 となる。式(5)の分子、分母は学習データ中の運動記号 λ と隠れ変数 s の共起回数の期待値、運動記号 λ の出現回数の期待値を表している。また、式(6)の分子、分母は、学習データ中の単語 ω と隠れ変数 s の共起回数の期待値、隠れ変数 s の出現回数の期待値を表している。

E-Step と M-Step の繰り返し計算によって、運動言語モデルの最適化を行う。

2.2 自然言語

自然言語処理に関して様々なモデルが提案されている。特に大規模な言語コーパスを扱う必要性のある自然言語処理では、統計

Fig. 3 Natural language model represents the sentence structure as transitions between two words. The node corresponds to the word, and the edge correspond to the transition from one word to another. The transition from the word ω_i to the word ω_j is represented by the conditional probability $P(\omega_j|\omega_i)$

的アプローチが有効であり、条件付確率場モデル (CRF: Conditional Random Field) [19] や HMM [20] を用いた日本語の形態素解析モデルが開発されている。本稿では、文章の構造を表現する自然言語モデルとして、最も構造が簡単な単語 bigram モデルを採用する。

Fig. 3 に自然言語モデルを示す。ノードは各単語を表し、エッジは単語間の遷移を表す。単語 bigram モデルでは、文章中の単語はその直前の単語にのみ依存するという仮定に基づき、単語間の関係性を単語 ω_i から単語 ω_j へ遷移する確率 $P(\omega_j|\omega_i)$ と、文章の先頭に単語 ω_i が現れる初期状態確率 π_{ω_i} により文章構造が表現される。

自然言語モデルの遷移確率と初期状態確率の最適値は以下のとおりに計算できる。

$$\pi_{\omega_i} = \frac{I(\omega_i)}{\sum_{j=1}^{N_{\omega}} I(\omega_j)} \quad (7)$$

$$P(\omega_j|\omega_i) = \frac{N(\omega_i, \omega_j)}{\sum_{k=1}^{N_{\omega}} N(\omega_i, \omega_k)} \quad (8)$$

ここで、 $I(\omega_i)$ は、学習データ中において文章の先頭に単語 ω_i が現れた回数、 $N(\omega_i, \omega_j)$ は単語 ω_i に単語 ω_j が続いた回数である。

3. 運動言語と自然言語の統合情報処理

運動記号から文章を生成する言語的解釈処理および文章から運動記号を生成する連想処理を運動言語モデルと自然言語モデルを用いた探索問題として計算する方法を提案する。本手法を用いることによってロボットの運動と言語間の相互推論計算が可能となる。

3.1 運動の言語的解釈

運動記号 λ から文章 (単語列) ω を生成する計算は、式 (9) で表される文章の探索問題とみなすことができる。ここで、文章 ω は l 個の単語の列 $(\omega[1], \omega[2], \dots, \omega[l])$ である。

$$\tilde{\omega} = \arg \max_{\omega} P(\omega|\lambda) \quad (9)$$

$$= \arg \max_{\omega} P(\omega[1], \dots, \omega[l]|\lambda) P(\omega|\omega[1], \dots, \omega[l])$$

$$(10)$$

$P(\omega[1], \dots, \omega[l]|\lambda)$ は運動の記号から単語の集合 $\{\omega[1], \omega[2], \dots, \omega[l]\}$ が生成される確率、 $P(\omega|\omega[1], \dots, \omega[l])$ は単語の集合 $\{\omega[1], \omega[2], \dots, \omega[l]\}$ から文章 ω が生成される確率を表している。各確率は運動言語モデルおよび自然言語モデルを用いて計算可能である。

単語の数や文章長の増加に伴い、式 (10) の探索空間は指数関数的な広がりを見せるため、効率的な探索方法が必要不可欠である。音響モデルと N-gram の単語モデルから成る音声認識処理では、音声入力から単語列を算出する方法として A*サーチアルゴリズムが用いられている [21]。本研究では、式 (10) の効率的な探索方法として、A*サーチの一種であるダイクストラ法を用いる。まず、対数尤度を用いて式 (10) を以下のように変形する。

$$\omega^o \approx \arg \max_{\omega} [\log P(\omega[1], \dots, \omega[l]|\lambda) + \log P(\omega|\omega[1], \dots, \omega[l])] \quad (11)$$

$$= \arg \max_{\omega} \left[\sum_{i=1}^l \log P(\omega[i]|\lambda) + \log \pi_{\omega[1]} + \sum_{i=2}^l \log P(\omega[i]|\omega[i-1]) \right] \quad (12)$$

k 番めまでの単語列 $(\omega[1], \omega[2], \dots, \omega[k])$ は決定されているが、 $k+1$ 番め以降の単語列 $(\omega[k+1], \omega[k+2], \dots, \omega[l])$ が未定の場合、 $\{P(\omega[i]|\lambda), P(\omega[i]|\omega[i-1]) : i = k+1, \dots, l\}$ の各確率は 1 以下であるという条件を用いて式 (11) の右辺の尤度を式 (13) (14) の右辺に示すように評価する。このような評価値を用いることによって探索空間が狭められ、かつ最適解を見つめられることが保障される。

$$\log P(\omega[1], \dots, \omega[l]|\lambda) \leq \sum_{i=1}^k \log P(\omega[i]|\lambda) \quad (13)$$

$$\log P(\omega|\omega[1], \dots, \omega[l]) \leq \log \pi_{\omega[1]} + \sum_{i=2}^k \log P(\omega[i]|\omega[i-1]) \quad (14)$$

運動記号から文章の生成は、この評価値が最大となる文章をダイクストラ法により見つけ出すことによって実用に耐えうる速度で計算可能となる。

3.2 言語から運動連想

文章 ω に相当する運動記号の連想情報処理は、式 (15) に記述されるように文章 ω から生成される尤度が最大となる運動記号 $\tilde{\lambda}$ を見つけ出す探索問題を解くことによって実現できる。

$$\tilde{\lambda} = \arg \max_{\lambda} P(\lambda|\omega) \quad (15)$$

$$= \arg \max_{\lambda} \frac{P(\lambda, \omega)}{P(\omega)} \quad (16)$$

$$= \arg \max_{\lambda} P(\lambda) P(\omega|\lambda) \quad (17)$$

$$= \arg \max_{\lambda} P(\omega|\lambda) \quad (18)$$

$$= \arg \max_{\lambda} P(\omega|\omega[1], \dots, \omega[l]) P(\omega[1], \dots, \omega[l]|\lambda) \quad (19)$$

Table 1 Training data of motions and manually given sentences

Motion
Sentences	学生が脚を組む (A student crosses his legs) 監督が脚を組む (A director crosses his legs) コーチが脚を組む (A coach crosses his legs)		
	主婦が料理する (A wife cooks) 主婦が包丁で切る (A wife cuts with the knife)		
Motion
Sentences	学生が土下座する (A student apologizes in Japanese style) 学生が謝る (A student apologizes)		
	主婦が箒で掃く (A wife sweeps with a broom) 主婦が掃除する (A wife cleans)		
Motion
Sentences	学生がテニスをする (A student plays tennis) 学生がテニスラケットを振る (A student swing a tennis racket) 選手がテニスをする (An athlete plays tennis)		
	選手が電話を掛ける (A student call on the phone) 学生が電話をする (A student makes a call)		
Motion
Sentences	選手がトスする (A player tosses a ball) 選手がバレーボールをする (An athlete play volleyball)		

$$= \arg \max_{\omega} \sum_{j=1}^l \log P(\omega[j]|\lambda) \quad (20)$$

ここで、文章 ω は単語列 $(\omega[1]\omega[2]\cdots\omega[l])$ として表現される。式 (16) は式 (15) からベイズ理論により導かれる。 $P(\omega)$ が運動記号 λ_i に依存しないという仮定に基づき、式 (16) から式 (17) へ変形される。式 (18) の導出では、各運動記号の事前確率 $P(\lambda_i)$ を等確率とする仮定を用いた。単語が生成される確率は運動記号にのみ依存するという仮定と対数尤度の計算を導入することによって、式 (20) が得られる。

展開済みの単語列を $(\omega[1]\omega[2]\cdots\omega[k])$ とすると、式 (20) の右辺の対数尤度項を式 (21) の右辺に示すように評価することによって、ダイクストラ法を適用することが可能となる。

$$\sum_{j=1}^l \log P(\omega[j]|\lambda) \leq \sum_{j=1}^k \log P(\omega[j]|\lambda) \quad (21)$$

このように、言語から運動記号の連想を探索問題として扱うことによって、高速に計算することが可能となる。

4. 実験結果

4.1 運動と言語の学習

本論文で提案する身体運動と自然言語を繋ぐ統計的情報処理

の枠組みの有効性をサンプリング時間 30 [ms] の 10 台のカメラから構成される光学式モーションキャプチャにて計測した人間の運動データを用いて検証した。Helen Hayes マーカーセットに従って被験者に取り付けられた 35 個のマーカーの位置から 34 自由度の人体骨格モデルを用いた逆運動学計算によって、ボディの水平方向の速度 v_x, v_y 、鉛直方向の位置 p_z 、ロール角 θ_x 、ピッチ角 θ_y 、ヨー角速度 $\dot{\theta}_z$ および、35 個のマーカーのボディ座標系における位置 ${}^B\mathbf{p}_i = ({}^Bp_{ix}, {}^Bp_{iy}, {}^Bp_{iz}) : (i = 1, 2, \dots, 35)$ を求める。これらを要素とする 111 次元の時系列信号を学習データとして各運動パターンに対応する HMM を最適化することによって運動記号を獲得する。本論文では、467 種類の運動記号を獲得した ($N_\lambda = 467$)。なお、学習データには 1 人の被験者の運動を計測したデータのみを用いた。これらの運動に対して複数の文章を説明文として付与した。付与された説明文は 764 個、全文章中で用いられた単語の種類は 241 個であった ($N = 764$, $N_\omega = 241$)。Table 1 に計測した運動と付与した説明文の一例を示す。一つの運動データは HMM を用いて一つの運動記号として認識される。また、その運動データに複数の文章が教師信号として与えられる。一つの運動記号に対する複数の文章が連想されるように学習データを用意した。このような学習データを用いて運動言語モデルおよび自然言語モデルの学習を行う。

Fig. 4 Sentences corresponding to each motion pattern are generated by the framework integrating the motion language model and the natural language model. Three sentences corresponding to the motion pattern are displayed in order of the likelihood that sentence is generated from the motion. The generated sentences are semantically and syntactically appropriate to the motion

なお、運動言語モデルの隠れ変数の数を 900 ($N_s = 900$)、学習での EM アルゴリズムの繰り返し計算は 10 回とした。本論文では、各運動パターンと文章を結び付ける枠組みを構築して

おり、運動パターン間の履歴、行為者や運動が作用する物体などの文脈情報を含んでいない。

[An athlete plays volleyball]

[A musician plays the violin]

[A student jumps forward]

Fig. 5 The motions are generated from three input sentences : “An athlete plays volleyball”, “A musician plays the violin”, “A student jumps forward”. The adequate motions are generated in response to the sentences through the framework integrating the motion language model and the natural language model

4.2 身体運動から言語の連想計算結果

光学式モーションキャプチャによって計測した人間の運動を運動記号として認識し、その運動記号から文章を連想する実験を行った。実験には学習と同じ被験者の400種類の運動を用いた。Fig. 4に計測した運動を人体骨格モデルに写像して得られる動作とその運動から連想された文章の一例が示されている。文章は、連想される尤度が高い3個の文章を解の候補として表示した。「脚を組む」運動から連想される尤度の高い文章は「学生が脚を組む」、「監督が脚を組む」、「コーチが脚を組む」である。これは、Table 1の学習データと比較して学習どおりに運動が言語として解釈されていることを示す。「土下座する」運動から連想される尤度の高い文章は「学生が謝る」「学生が土下座する」であり、学習した文章が連想されている。また、第3番めの文章候補として「選手が謝る」が連想されている。この「選手が謝る」という文章は学習データには存在しない。しかし、学習データ中で他の運動に対して「選手」と「学生」という単語が類似した文脈で頻繁に使われているため、主語の候補として「選手」を当てはめた文章を連想している。「ピアノを演奏する」運動では、動作が働き掛ける物体である「ピアノ」は

計測されていないが、運動から「ピアノ」という単語を連想し、「音楽家がピアノを弾く」と学習どおりに文章を連想している。他の運動に関する文章を連想する実験結果からも、第1番めの文章候補として出力されている文章は運動を適切に表現していることが確認できる。

400種類の運動に文章があらかじめ付与されている。この文章を正解とした場合、運動から連想された文章と正解文章の一致する割合を調査した。連想された第1番めの文章候補が正解文章と一致した運動は167種類であった。第1, 2番めの文章候補のどちらか一方でも正解文章と一致する運動は293種類、第1, 2, 3番めの文章候補のどれか一つでも正解文章と一致する運動は353種類であった。すなわち、運動から連想される文章の正解率は0.42, 0.73, 0.88であった。

このように、運動言語モデルが運動記号から動詞だけでなく関係の深い名詞を生成していること、自然言語モデルが連想された単語集合から適切な単語の並びをした文章を生成していること、運動から言語の連想計算を運動記号から文章の探索問題として解く手法が妥当であることが実験によって示せた。

4.3 言語から身体運動の生成計算

文章から運動記号を連想して運動を生成する実験を行った。入力された文章を構成する単語の集合を形態素解析「茶筌」によって抽出する。得られた単語集合から連想される尤度が最大の運動記号を求める。この運動記号から、ボディの水平方向の速度 v_x, v_y , 鉛直方向の位置 p_z , ロール角 θ_x , ピッチ角 θ_y , ヨー角速度 $\dot{\theta}_z$ および, 35 個のマーカのボディ座標系における位置 ${}^B\mathbf{p}_i = ({}^Bp_{ix}, {}^Bp_{iy}, {}^Bp_{iz}) : (i = 1, 2, \dots, 35)$ の時系列信号を復元し, 35 個のマーカの位置から逆運動学計算によってロボットの関節角度を計算する。この一連の計算によって文章から運動が生成可能となる。

学習データに含まれる文章から 400 個の文章を無作為に選択し, その文章から運動を生成する実験を行った。Fig. 5 は言語からの運動生成の実験結果の一例を表す。「選手がバレーボールをする」という文章から「トスを上げる」運動が生成されている。同様に「音楽家がバイオリンを弾く」および「学生が前にジャンプする」という文章から「バイオリンを演奏する」, 「ジャンプする」運動が生成されている。また, 学習データの中から入力された文章が教師として与えられている運動を選び出し, その運動の認識結果として得られる運動記号と文章から連想された運動記号の一致する割合を調べた。その結果, 289 個の文章で一致すること (一致率 0.72) を確認した。このように入力された文章から適切に運動が生成されており, 提案する言語から単語の集合を経て運動記号を連想する計算方法が有効に働くことを確認した。

5. 結 言

本研究で得られた結論は以下のとおりである。

- (1) 運動の記号と単語の連想構造を表現する運動言語モデルと文章構造を統計的に表現する自然言語モデルを統合する枠組みを提案した。運動言語モデルは, 運動記号, 隠れ変数, 単語の 3 層構造の統計モデルであり, 自然言語モデルは 2 単語を一つのまとまりとして捉えた bigram モデルを用いた。
- (2) 運動を言語として解釈する計算法を提案した。運動言語モデルによって運動記号から単語の集合を連想する。その単語の並び方を自然言語モデルを通じて評価し, 適切な文章を出力することで実現される。また, 言語から運動を生成する計算方法を提案した。文章を構成する単語の集合を抽出し, その単語から連想される運動記号を運動言語モデルを通じて求めることで実現した。これらの運動と言語間の双方向変換は, 探索問題として表現され高速に計算が可能である。
- (3) 467 個の運動と 764 個の文章 (用いられた単語の種類は 241 個) の学習データから運動言語モデルと自然言語モデルを構築した。光学式モーションキャプチャシステムにて計測した運動から文章を連想する実験を行った。運動言語モデルは運動から適切な単語を連想し, 自然言語モデルは適切な単語の並びをした文章を生成する機能を果たすことによって, 運動に対して正しい文章が作り出されることを確認した。
- (4) 文章から運動を生成する実験を行った。文章を構成する単語から適切な運動記号を連想し, 運動記号から実際のロボッ

トの身体運動を生成した。提案する枠組みによってロボットが入力された文章を実世界の運動情報へと適切に変換することが可能であることを実験によって確認した。

謝 辞 本研究は, 平成 23 年度独立行政法人科学技術振興機構戦略的創造研究推進事業 (さきがけ) 「行動の記号化を基盤とした身振り・言語を通じてコミュニケーションするロボットの知能設計」 (代表: 高野渉) の支援を受けた。

参 考 文 献

- [1] F.D. Saussure: Course in general linguistics. McGraw-Hill Book Company, 1966.
- [2] T. Deacon: Symbolic species: The co-evolution of language and the brain. W.W.Norton and Company Inc, 1997.
- [3] G. Rizzolatti, L. Fogassi and V. Gallese: "Neurophysiological mechanisms underlying the understanding and imitation of action," Nature Reviews, vol.2, pp.661-670, 2001.
- [4] V. Gallese and A. Goldman: "Mirror neuron and the simulation theory of mind-reading," Trends in Cognitive Sciences, vol.2, no.12, pp.493-501, 1998.
- [5] M. Haruno, D. Wolpert and M. Kawato: "Mosaic model for sensorimotor learning and control," Neural Computation, vol.13, pp.2201-2220, 2001.
- [6] J. Tani and M. Ito: "Self-organization of behavioral primitives as multiple attractor dynamics: A robot experiment," IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans, vol.33, no.4, pp.481-488, 2003.
- [7] H. Kadone and Y. Nakamura: "Symbolic memory for humanoid robots using hierarchical bifurcations of attractors in nonmonotonic neural networks," Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.2900-2905, 2005.
- [8] A. Billard and R. Siegwart: "Robot learning from demonstration," Robotics and Autonomous Systems, vol.47, pp.65-67, 2004.
- [9] T. Inamura, I. Toshima, H. Tanie and Y. Nakamura: "Embodied symbol emergence based on mimesis theory," International Journal of Robotics Research, vol.23, no.4, pp.363-377, 2004.
- [10] T. Nakamura, T. Nagai and N. Iwahashi: "Multimodal object categorization by a robot," Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.2415-2420, 2007.
- [11] Y. Sugita and J. Tani: "Learning semantic combinatoriality from the interaction between linguistic and behavioral processes," Adaptive Behavior, vol.13, no.1, pp.33-52, 2005.
- [12] T. Ogata, M. Murase, J. Tani, K. Komatani and H.G. Okuno: "Two-way translation of compound sentences and arm motions by recurrent neural networks," Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1858-1863, 2007.
- [13] K. Sugiura, N. Iwahashi, H. Kashioka and S. Nakamura: "Active learning of confidence measure function in robot language acquisition framework," Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1774-1779, 2010.
- [14] T. Nakamura, T. Nagai and N. Iwahashi: "Bag of multimodal lda models for concept formation," Proceedings of the IEEE International Conference on Robotics and Automation, pp.6233-6238, 2011.
- [15] O. Arikan, D.A. Forsyth and J.F. O'Brien: "Motion synthesis from annotations," ACM Transactions on Graphics, vol.22, no.3, pp.402-408, 2003.
- [16] C. Rose, B. Bodenheimer and M.F. Cohen: "Verbs and adverbs: Multidimensional motion interpolation," IEEE Computer Graphics and Application, vol.18, no.5, pp.32-40, 1998.

- [17] W. Takano, K. Yamane and Y. Nakamura: "Capture database through symbolization, recognition and generation of motion patterns," Proceedings of the IEEE International Conference on Robotics and Automation, pp.3092–3097, 2007.
- [18] W. Takano, D. Kulic and Y. Nakamura: "Interactive topology formation of linguistic space and motion space," Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1416–1422, 2007.
- [19] T. Kudo, K. Yamamoto and Y. Matsumoto: "Applying conditional random fields to Japanese morphological analysis," Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pp.230–237, 2004.
- [20] K. Takeuchi and Y. Matsumoto: "HMM parameter learning for Japanese morphological analyzer," Proceedings of the 10th Pacific Asia Conference on Language, Information and Computation, pp.163–172, 1995.
- [21] A. Lee, T. Kawahara and S. Doshita: "Large vocabulary continuous speech recognition parser based on A* search using grammar category-pair constraint," Journal of information processing, vol.40, no.4, pp.1374–1382, 1999.

付録 A. 運動言語モデルの最適化

式 (1)～(3) を用いて表される評価関数 Φ に関して, Jensen の不等式から式 (A.1) が得られる.

$$\begin{aligned}
 \Phi(\theta^{[t]}) &= \sum_{k=1}^N \log \left[\prod_{i=1}^{n_k} \sum_{j=1}^{N_s} P(s_j | \lambda^k, \omega_i^k) \frac{P(\omega_i^k, s_j | \lambda^k, \theta^{[t]})}{P(s_j | \lambda^k, \omega_i^k)} \right] \\
 &= \sum_{k=1}^N \log \left\{ \prod_{i=1}^{n_k} E_{P(s | \lambda^k, \omega_i^k)} \left[\frac{P(\omega_i^k, s | \lambda^k, \theta^{[t]})}{P(s | \lambda^k, \omega_i^k)} \right] \right\} \\
 &= \sum_{k=1}^N \sum_{i=1}^{n_k} \log E_{P(s | \lambda^k, \omega_i^k)} \left[\frac{P(\omega_i^k, s | \lambda^k, \theta^{[t]})}{P(s | \lambda^k, \omega_i^k)} \right] \\
 &\geq \sum_{k=1}^N \sum_{i=1}^{n_k} E_{P(s | \lambda^k, \omega_i^k)} \left[\log \frac{P(\omega_i^k, s | \lambda^k, \theta^{[t]})}{P(s | \lambda^k, \omega_i^k)} \right] \\
 &\equiv \mathcal{F}(P(s | \lambda, \omega), \theta^{[t]}) \quad (\text{A.1})
 \end{aligned}$$

ここで, $\theta^{[t]}$ は t 回目の最適化計算によって推定されたモデルパラメータの集合であり, $E_q[p]$ は分布 q の下での p の期待値を表している. $\mathcal{F}(P(s | \lambda, \omega), \theta^{[t]})$ は評価関数の下限値に相当する. 評価関数 $\Phi(\theta^{[t]})$ と評価関数の下限値 $\mathcal{F}(P(s | \lambda, \omega), \theta^{[t]})$ の差分はベイズの理論より式 (A.2) で表される.

$$\begin{aligned}
 \Phi(\theta^{[t]}) - \mathcal{F}(P(s | \lambda, \omega), \theta^{[t]}) &= \sum_{k=1}^N \sum_{i=1}^{n_k} \left\{ \log P(\omega_i^k | \lambda^k, \theta^{[t]}) \right. \\
 &\quad \left. - E_{P(s | \lambda^k, \omega_i^k)} \left[\log \frac{P(\omega_i^k, s | \lambda^k, \theta^{[t]})}{P(s | \lambda^k, \omega_i^k)} \right] \right\} \\
 &= \sum_{k=1}^N \sum_{i=1}^{n_k} \left\{ \log P(\omega_i^k | \lambda^k, \theta^{[t]}) \right. \\
 &\quad \left. - E_{P(s | \lambda^k, \omega_i^k)} \left[\log \frac{P(\omega_i^k | \lambda^k, \theta^{[t]}) P(s | \lambda^k, \omega_i^k, \theta^{[t]})}{P(s | \lambda^k, \omega_i^k)} \right] \right\} \\
 &= \sum_{k=1}^N \sum_{i=1}^{n_k} E_{P(s | \lambda^k, \omega_i^k)} \left[\log \frac{P(s | \lambda^k, \omega_i^k)}{P(s | \lambda^k, \omega_i^k, \theta^{[t]})} \right]
 \end{aligned}$$

$$= \sum_{k=1}^N \sum_{i=1}^{n_k} KL \left(P(s | \lambda^k, \omega_i^k) \| P(s | \lambda^k, \omega_i^k, \theta^{[t]}) \right) \quad (\text{A.2})$$

式 (A.2) はモデルでの隠れ変数の分布 $P(s | \lambda^k, \omega_i^k, \theta^{[t]})$ と隠れ変数の分布 $P(s | \lambda^k, \omega_i^k)$ のカルバック距離を表す. カルバック距離は

$$KL \left(P(s | \lambda^k, \omega_i^k) \| P(s | \lambda^k, \omega_i^k, \theta^{[t]}) \right) \geq 0 \quad (\text{A.3})$$

の関係を満たし, $P(s | \lambda^k, \omega_i^k) = P(s | \lambda^k, \omega_i^k, \theta^{[t]})$ の時, 式 (A.3) の等号が成立する. 隠れ変数の分布 $P(s | \lambda^k, \omega_i^k)$ をモデルでの隠れ変数の分布 $P(s | \lambda^k, \omega_i^k, \theta^{[t]})$ として推定すること (E-Step) によって式 (4) が得られる.

式 (A.2) より評価関数の変化量 $\Delta\Phi$ は式 (A.4) として表される.

$$\begin{aligned}
 \Delta\Phi &= \Phi(\theta^{[t+1]}) - \Phi(\theta^{[t]}) \\
 &= \mathcal{F}(P(s | \lambda, \omega), \theta^{[t+1]}) - \mathcal{F}(P(s | \lambda, \omega), \theta^{[t]}) \\
 &\quad + \sum_{k=1}^N \sum_{i=1}^{n_k} KL \left(P(s | \lambda^k, \omega_i^k) \| P(s | \lambda^k, \omega_i^k, \theta^{[t+1]}) \right) \\
 &\quad - \sum_{k=1}^N \sum_{i=1}^{n_k} KL \left(P(s | \lambda^k, \omega_i^k) \| P(s | \lambda^k, \omega_i^k, \theta^{[t]}) \right) \quad (\text{A.4})
 \end{aligned}$$

ここで, 隠れ変数の分布 $P(s | \lambda^k, \omega_i^k)$ をモデルでの隠れ変数の分布 $P(s | \lambda^k, \omega_i^k, \theta^{[t]})$ とすることによって, 式 (A.4) の第3項～第4項は0以上となる. したがって, 評価関数の変化量 $\Delta\Phi$ を正にするには, $\mathcal{F}(P(s | \lambda, \omega), \theta^{[t+1]})$ を最大化すればよいことになる. 評価関数の下限値 $\mathcal{F}(P(s | \lambda, \omega), \theta^{[t+1]})$ を最大化するモデルパラメータを求める (M-Step) 繰り返し計算によってモデルを最適化することができる. 式 (A.5) に評価関数の下限値を最大化する問題を示す.

$$\begin{aligned}
 \arg \max_{\theta^{[t+1]}} \mathcal{F}(P(s | \lambda, \omega), \theta^{[t+1]}) &= \arg \max_{\theta^{[t+1]}} \sum_{k=1}^N \sum_{i=1}^{n_k} \sum_{j=1}^{N_s} P(s_j | \lambda^k, \omega_i^k) \\
 &\quad \times \left[\log P(\omega_i^k | s_j, \theta^{[t+1]}) P(s_j | \lambda^k, \theta^{[t+1]}) \right. \\
 &\quad \left. - \log P(s_j | \lambda^k, \omega_i^k) \right] \\
 &= \arg \max_{\theta^{[t+1]}} \sum_{k=1}^N \sum_{i=1}^{n_k} \sum_{j=1}^{N_s} P(s_j | \lambda^k, \omega_i^k) \\
 &\quad \times \log P(\omega_i^k | s_j, \theta^{[t+1]}) P(s_j | \lambda^k, \theta^{[t+1]}) \quad (\text{A.5})
 \end{aligned}$$

ここで, 第2項の $P(s_j | \lambda^k, \omega_i^k)$ はモデルパラメータ $\theta^{[t+1]}$ に依存しないため省略した. 拘束条件 $\left(\sum_{i=1}^{N_s} P(s_i | \lambda, \theta^{[t+1]}) = 1, \sum_{i=1}^{N_\omega} P(\omega_i | s, \theta^{[t+1]}) = 1 \right)$ 付き最大化問題なので, Lagrange の未定乗数法を適用して式 (A.5) の評価関数を式 (A.6) とする.

$$\begin{aligned} \mathcal{L} = & \sum_{k=1}^N \sum_{i=1}^{n_k} \sum_{j=1}^{N_s} P(s_j | \lambda^k, \omega_i^k) \\ & \times \left[\log P(\omega_i^k | s_j, \theta^{[t+1]}) + \log P(s_j | \lambda^k, \theta^{[t+1]}) \right] \\ & - \sum_{i=1}^{N_\lambda} \alpha_i \left(\sum_{j=1}^{N_s} P(s_j | \lambda_i, \theta^{[t+1]}) - 1 \right) \\ & - \sum_{i=1}^{N_s} \beta_i \left(\sum_{j=1}^{N_\omega} P(\omega_j | s_i, \theta^{[t+1]}) - 1 \right) \end{aligned} \quad (\text{A.6})$$

式 (A.6) を各モデルパラメータで偏微分した値を 0 とおいて解くことによって、運動言語モデルの最適化計算を行う。

$$\frac{\partial \mathcal{L}}{\partial P(s | \lambda, \theta^{[t+1]})} = 0 \quad (\text{A.7})$$

$$\frac{\partial \mathcal{L}}{\partial P(\omega | s, \theta^{[t+1]})} = 0 \quad (\text{A.8})$$

式 (A.7) (A.8) は解析的に解くことが可能であり、局所最適なモデルパラメータは式 (5) (6) として求めることができる。

高野 渉 (Wataru Takano)

1999 年京都大学工学部物理工学科卒業。2001 年同大学大学院工学研究科精密工学専攻修士課程修了。同年三菱自動車工業株式会社入社。2006 年東京大学大学院情報理工学系研究科知能機械情報学専攻修士課程修了。2006 年同大学特任助教、2007 年同大学助教、2009 年同大学講師、現在に至る。博士 (情報理工学)。ヒューマノイドロボットの統計的脳型情報処理、自動車の知能に関する研究に従事。(日本ロボット学会正会員)

鮎澤 光 (Ko Ayusawa)

2006 年 3 月東京大学工学部機械情報工学科卒業。2011 年 3 月同大学大学院情報理工学系研究科知能機械情報学専攻博士課程修了。現在同大学大学院情報理工学系研究科知能機械情報学専攻特任研究員。博士 (情報理工学)。ヒューマノイドの運動学・動力学・力学同定・制御、ヒトの運動解析・質量パラメータの同定・筋張力計算に関する研究に従事。(日本ロボット学会正会員)

濱野聖也 (Seiya Hamano)

2010 年 3 月東京大学工学部機械情報工学科卒業。2012 年 3 月同大学大学院情報理工学系研究科知能機械情報学専攻修士課程修了。現在ヤフー株式会社勤務。大学院在学時はヒューマノイドロボットの知能に関する研究に従事。

(日本ロボット学会学生会員)

梅澤慶介 (Keisuke Umezawa)

2012 年東京大学工学部機械情報工学科卒業。現在同大学大学院情報理工学系研究科知能機械情報学専攻修士課程在学中。ヒューマノイドロボットの知能に関する研究に従事。(日本ロボット学会学生会員)

中村仁彦 (Yoshihiko Nakamura)

1977 年京都大学工学部精密工学科卒業。1982 年同大学大学院博士課程退学、1982 年同大学助手。1987 年カリフォルニア大学サンタバーバラ校助教授、準教授。1991 年東京大学工学部機械情報工学科助教授。現在同大学大学院情報理工学系研究科教授。工学博士。非ホロノミックロボット、脳型情報処理、ヒューマノイドや CG の力学計算、外科手術用ロボットなどの研究に従事。システム制御情報学会、計測自動制御学会、日本コンピュータ外科学会、IEEE、ASME などの会員。日本機械学会フェロー、日本 IFToMM 会議委員長、IEEE Robotics and Automation Society, Distinguished Lecturer (2006~2007)。(日本ロボット学会正会員)