3B1-OS22b-6

運動と自然言語の統計的推論を用いた運動データベースの設計

Design of the Motion Database from Integration of Motion Symbols and Natural Language

高野　渉[*1]　　中村　仁彦[*1]
Wataru Takano　Yoshihiko Nakamura

[*1]東京大学
The University of Tokyo

We have been developing intelligent humanoid robots based on symbolization of motion patterns. The symbolic framework makes it possible for the robots to recognize observation and generate their behaviors. Moreover, the integration of the symbols and natural language will realize the cohabitation with human and robots, which can efficiently and effectively communicate and cooperate through linguistic inference. This paper proposes a novel approach to design motion database, under which stochastic associative computation between motion symbols and natural language lies. We implemented this database into small humanoid robots, which can understand observation as language, and generate motion patterns corresponding to input speech. We can also retrieve necessary motion data from inputs of key words or sentences by using the motion database. This technology can be applied in various kinds of fields, robotics, rehabilitation, sports engineering and animation.

1. はじめに

実世界の多様な事象を記号として理解し，記号の組み合わせから無限の事象を創造する．この記号の抽象化と生成性が人間の知能の本質である．特に，言語は同じ社会・文化に属する人間が同じ事象を共有・伝達する記号システムである．人間と共存するロボットにもこのような言語・記号システムが求められつつある．

動作をモデルパラメータとして学習する枠組み[Inamura 04][Sugita 05][Sugiura 10]，画像・触覚などの物体のマルチモーダル情報から概念記号を獲得する手法など実世界の事象を記号として獲得するロボットの研究が行われてきた[Nakamura 11]．筆者らも，人間の全身運動を記号化し，記号を用いて人間の動作を理解しながら人間に身振りを使って働き掛けるヒューマノイドロボットの知能を構築してきた[Takano 06]．多様な動きを理解・生成するには，膨大な数の運動の記号化と言語を繋げる数理モデルが必要不可欠である．

本論文では，運動の記号と自然言語を結ぶ付ける計算論およびそれを実装したヒューマノイドロボットの実験結果について述べる．さらに，これを基盤とした開発中の運動データベースについても紹介する．

2. 運動の記号と自然言語の統合

運動を隠れマルコフモデルによって学習して獲得される運動の記号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．図 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．

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．各確率は，隠れ変数の分布を推定する E-step

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

および，訓練データが生成される確率を最大化する M-step

$$P(s|\lambda)=\frac{\sum_{k=1}^{N}\sum_{i=1}^{n_k}\delta(\lambda,\lambda^k)P(s|\lambda^k,\omega_i^k)}{\sum_{j=1}^{N_s}\sum_{k=1}^{N}\sum_{i=1}^{n_k}\delta(\lambda,\lambda^k)P(s_j|\lambda^k,\omega_i^k)} \tag{2}$$

$$P(\omega|s)=\frac{\sum_{k=1}^{N}\sum_{i=1}^{n_k}\delta(\omega,\omega_i^k)P(s|\lambda^k,\omega_i^k)}{\sum_{j=1}^{N_\omega}\sum_{k=1}^{N}\sum_{i=1}^{n_k}\delta(\omega_j,\omega_i^k)P(s|\lambda^k,\omega_i^k)} \tag{3}$$

を繰り返す EM アルゴリズムによって最適化される．ここで，運動記号，単語，隠れ状態の集合は，各々$\{\lambda_i|i=1,2,\cdots,N_\lambda\}$，$\{\omega_i|i=1,2,\cdots,N_\omega\}$，$\{s_i|i=1,2,\cdots,N_s\}$ であり，運動記号 λ^k は k 番目に学習された運動を表す運動記号，$\left(\omega_1^k\omega_2^k\cdots\omega_{n_k}^k\right)$ はその運動を表す文章（単語列）である．また，θ は，運動言語モデルのパラメータ集合である．

この運動言語モデルによって，ある運動記号から単語の連想確率が

$$P(\omega|\lambda)=\sum_{i=1}^{N_s}P(\omega|s_i)P(s_i|\lambda) \tag{4}$$

として計算することができる．自然言語モデルは，単語の出現はその直前の $N-1$ 個の単語の並びにのみ依存するとする Ngram モデルを用いた．すなわち，単語 ω の出現確率は，その直前の単語列 $\boldsymbol{\omega_1^{N-1}}$ を条件とする

$$P(\omega|\boldsymbol{\omega_1^{N-1}})=\frac{\mathcal{N}(\boldsymbol{\omega_1^{N-1}},\omega)}{\mathcal{N}(\boldsymbol{\omega_1^{N-1}})} \tag{5}$$

連絡先: 高野渉，東京大学，東京都文京区本郷 7-3-1 東京大学工学部，takano@ynl.t.u-tokyo.ac.jp

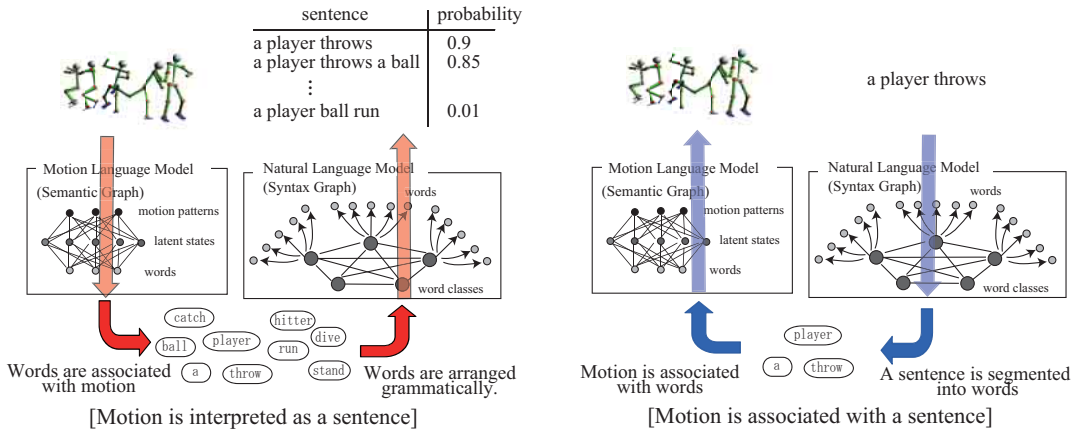

図 1: Overview of integration of a motion language model with a natural language model. The motion language model represents relationship among motion symbols and morpheme words via latent variables as a graph structure, where nodes on 1st, 2nd and 3rd layer indicate the motion symbols, the morpheme words and the latent variables respectively. The natural language model represents the dynamics of language which means the order of words in sentences. The motion language model and the natural language model are equivalent to semantics and syntax. By integrating two functions, linguistic processing for robots can be realized.

	mid	file	description0	description1	description2
	1	20100916_takano/bow1.trc	学生がお辞儀する	投手がお辞儀する	選手がお辞儀する
	4	20100916_takano/catch1.trc	選手が捕球する		
	7	20100916_takano/clap1.trc	選手が拍手する	学生が拍手する	
	10	20100916_takano/crouch1.trc	選手が屈む	投手が屈む	投手がマウンドで屈む
	13	20100916_takano/dodish1.trc	主婦が洗う	主婦が食器を洗う	主婦が食器洗いをする
	16	20100916_takano/drink1.trc	選手が飲む	学生が飲む	
	19	20100916_takano/folderm1.trc	選手が腕組みする		
	22	20100916_takano/highkick_left1.trc	格闘家がキックする	選手がキックする	格闘家が左足でキックする
	25	20100916_takano/highkick_right1.trc	格闘家がキックする	選手がキックする	格闘家が右足でキックする
	30	20100916_takano/holdup3.trc	選手が持ち上げる	選手が仲間を持ち上げる	

図 2: Database consists of 537 motion data, each of which is given description labels. The database can be used for retreival of motion data from words.

として求められる．ここで，$\mathcal{N}(\boldsymbol{\omega}_1^{N-1})$，$\mathcal{N}(\boldsymbol{\omega}_1^{N-1},\omega)$ は，学習データ中に単語列 $\boldsymbol{\omega}_1^{N-1}$ および $\left\{\boldsymbol{\omega}_1^{N-1},\omega\right\}$ が現れた回数である．

この運動言語モデルと自然言語モデルを用いることによって，運動記号から単語列，および単語列から運動記号への変換は確率が変換される確率が最大となる解を求める探索問題として解くことができる．

3. 実験結果

人間の運動データを光学式モーションキャプチャシステムに計測した．本研究では，537 個の運動データから構成される運動データベースを構築した．Fig.2 に作成した運動データベースの表を示す．データベースには，運動データとその運動を表現する説明文，被験者，計測日および動画に関する情報が格納されている．リレーショナルデータベース管理システムの一種である SQL によって運動データベースは構築されており，単語や文章から必要な運動データを検索することも可能となっている．

運動データベースを用いて観察した運動を言語として解釈する実験を行った．表 1 に各運動から生成される 5 つの文章，各文章を構成する単語の集合が運動言語モデルから生成される対数尤度 $\log P(\omega_1^*,\cdots,\omega_{n_*}^*|\lambda)$，各文章が自然言語モデルから生成される対数尤度 $\log P(\boldsymbol{\omega}|\omega_1^*,\cdots,\omega_{n_*}^*)$，および運動記号から各文章が生成される対数尤度 $\log P(\ \boldsymbol{\omega}|\lambda^o)$ を示す．「走る」運動データに対して，文章「選手　が　走る」,「打者　が　走る」と学習データに用いた文章が生成される．「握手する」運動データに対しても,「選手　が　握手する」,「打者　が　握手する」と適切な文章が生成される．他の「投げる」,「手を広げる」といった運動データに対応する適切な文章が生成されていることが示されている．

また，運動データベースを小型ヒューマノイドロボットに実装した．このロボットに発話入力を与えたときに生成される運動を検証した．入力された発話は音声認識ソフトウェア Julius によって文章テキストに変換される．この文章から自然言語モデル・運動言語モデルを通じて運動記号が選択される．文章から想起される確率の高い運動記号を 4 つ求め，各運動記号に相当する運動を小型ヒューマノイドロボットが行う．Fig.3 の右図は,「主婦が掃除をする」の発話を聞いたときに生成される運動をロボットが行っている実験結果である．「掃除機をかける」運動が生成されている．Fig.3 の左図は,「学生が歩く」の発話を聞いたときに生成される確率の高い運動をロボットが行っている実験結果である．最も確率の高い運動として「歩く」動作が生成されている．2 番目および 4 番目の候補として「お辞儀する」「読書する」運動が生成された．これは，学習データとして「学生がお辞儀する」「学生が読書する」という文章と「お辞儀する」「読書する」運動記号の組み合わせが与えられたため,「学生」から「お辞儀する」「読書する」の運動記号が想起されたためである．このように，文章中の動詞のみだけでなく名詞からも動きを連想することが提案するモデルの特徴でもある．

4. 結言

運動の記号から単語を連想する運動言語モデル，単語の並びを表現する自然言語モデル，これら 2 つのモデルを統合ず

表 1: Experimental results. This table shows sentences which motion patterns are interpreted as.

motion pattern	sentences	$\log P(\omega_1^*, \cdots, \omega_{n_*}^* \mid \lambda^o)$	$\log P(\boldsymbol{\omega} \mid \omega_1^*, \cdots, \omega_{n_*}^*)$	$\log P(\boldsymbol{\omega} \mid \lambda^o)$
run	a player runs	-7.53	-12.74	-20.28
	a hitter runs	-7.53	-14.60	-22.14
	a runner runs	-7.53	-14.87	-22.40
	a player a player	-8.63	-13.81	-22.44
	a player a hitter	-8.63	-15.67	-24.40
shake a hand	a player shakes a hand	-7.47	-15.41	-22.88
	a hitter shakes a hand	-7.69	-17.26	-24.96
	a runner shakes a hand	-7.47	-17.53	-25.00
	a player a hitter	-9.35	-15.67	-25.02
	a hitter a player	-9.35	-15.67	-25.02
throw	a player throws	-8.03	-13.49	-21.52
	a pitcher throws	-7.34	-14.49	-21.83
	throw a pitcher	-8.44	-13.63	-22.07
	a ball throws	-8.03	-15.03	-23.07
	throw a ball	-9.13	-14.17	-23.30
open arms	a player opens his arms	-13.17	-21.56	-34.74
	his arms open a player	-13.17	-21.56	-34.74
	a pitcher opens his arm	-13.17	-21.64	-34.82
	his arm, a pitcher opens	-13.17	-21.64	-34.84
	a player a player his arm	-14.97	-21.24	-36.21

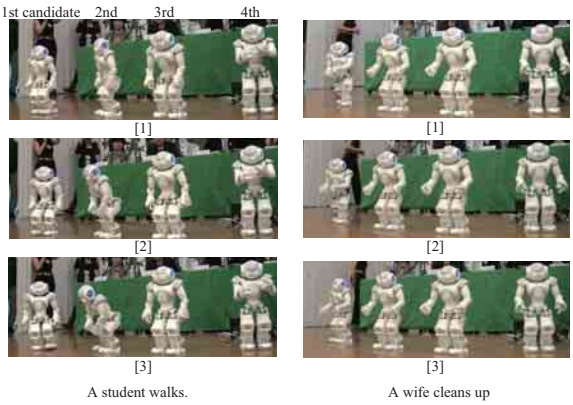

図 3: Our proposed system associate motion symbols from utterance. The utterance is converted to a sentence through the speech recognition software of Julius. The sentence is converted to 4 motion symbols, which are the most likely to be generated from the sentence. The small robots perform motion patterns corresponding to the generated motion syhmbols respectively.

る計算方法を提案した．ロボットが人間の運動を観察し，それを言語として解釈できることを実験によって示した．また，小型ヒューマノイドロボットに実装し，入力された音声から適切に行動が連想されることを示した．

なお，本研究は，平成２２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）の支援を受けて行った．

参考文献

[Inamura 04] Inamura, T., Toshima, I., Tanie, H., and Nakamura, Y.: Embodied symbol emergence based on mimesis theory, *International Journal of Robotics Research*, Vol. 23, No. 4, pp. 363–377 (2004)

[Nakamura 11] Nakamura, T., Nagai, T., and Iwahashi, N.: Bag of multimodal LDA Models for Concept Formation, in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 6233–6238 (2011)

[Sugita 05] Sugita, Y. and Tani, J.: Learning semantic combinatoriality from the interaction between linguistic and behavioral processes, *Adaptive Behavior*, pp. 33–52 (2005)

[Sugiura 10] Sugiura, K., Iwahashi, N., Kashioka, H., and Nakamura, S.: Active Learning of Confidence Measure Function in Robot Language Acquisition Framework, in *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1774–1779 (2010)

[Takano 06] Takano, W., Yamane, K., Sugihara, T., Yamamoto, K., and Nakamura, Y.: Primitive Communication based on Motion Recognition and Generation with Hierarchical Mimesis Model, in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3602–2609 (2006)