解　説

ロボティクスと自然言語処理

Natural Language Processing in Robotics

持　橋　大　地*　*統計数理研究所
Daichi Mochihashi*　*The Institute of Statistical Mathematics

1. はじめに

筆者は自然言語処理の研究者であるが, 2014 年頃から, 科研費基盤 (B)「潜在意味空間において感覚情報を言語化し言語的思考を行うロボットの実現」の分担者, 2018 年から同基盤 (B) 「運動の統計的理解と動力学に基づく適応的確率ロボティクス」の代表者として, 本特集の執筆者でもある知能ロボティクスの研究者と共同研究を行ってきた. そうした中で, ロボティクスの国際会議である IROS に参加したり [1] [2], 機会をいただいて 2019 年の CoRL ではチュートリアルを行ったりしている [3].

まずは運動の制御といった物理面が重要なロボティクスではあるが, 今後はロボティクスの物理的側面が成熟するのに伴い, その上で動く情報処理がより重要になってくると考えられる. 特に, ロボットが人間と共生して動作することを考える場合, ロボットとの相互の情報伝達の手段としては, 複雑で構造的な知識を伝えることのできる自然言語によらざるを得ない. ゆえに, ロボットにどのように自然言語を理解させるのか (人間→ロボット, **言語理解**), およびロボットが自分の状態や意図をどのように人間に伝えるのか (ロボット→人間, **言語生成**) の問題は今後ますます重要性を増すと考えられる. そこで本稿では, 一般のロボティクス研究者・技術者に向けて, ロボティクスに関係する自然言語処理の概観, 現状および課題を, 筆者の共同研究の成果を交じえつつ紹介する.

2. 自然言語処理とロボティクス

自然言語処理は計算言語学ともいわれる. コンピュータサイエンスの一分野であるため, ジャーナルより国際会議の論文が大きな意味を持つ分野であるが, トップ国際会議である ACL (Association of Computational Linguistics) では, 最近になって論文の投稿エリアとして「ロボティクス」が明示されるようになっている. 例えば, 最近の ACL-IJCNLP 2021 においては, “Language Grounding to Vision, Robotics and Beyond” が存在しており, 論文の査読はこのエリアの中で行われるため, ロボティクスとの接続は明確に重要なエリアとして認識されているといえる.

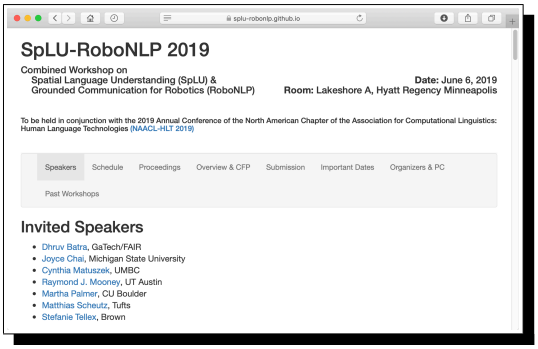


図 1　自然言語処理のトップ国際会議の 1 つである, NAACL 2019 における言語ロボティクスと空間認識のワークショップ SpLU-RoboNLP 2019 のホームページ.

実際に, 2017 年の ACL ではワークショップとして “Language Grounding for Robotics” † が企画されており, それを受けて 2019 年には図 1 のように NAACL において SpLU-RoboNLP 2019 (Spatial Language Understanding & Grounded Communication for Robotics) †† が開催されて, それぞれ 13 本, 9 本の論文が採録されている†††.

こうしてロボティクスにおいても自然言語処理の重要性が増し, また言語処理側でもロボット上での言語処理が視野に入り始めていることから, 谷口 (立命館大) と筆者らは 2018 年に計測自動制御学会の研究会として LangRobo 研究会‡を立ち上げ, 6 回にわたって研究会を開いて, 言語ロボティクスに関する議論を重ねてきた. この議論の内容は, 2019 年 2 月に開かれた NII 湘南会議 “Learning to communicate: Challenges in language learning by AI, robots and humans” を経て, 幹事団を中心としたロボティクス, 自然言語処理, 言語学にわたるメンバー (谷口忠大 (立命館), 持橋大地 (統数研), 長井隆行 (電通大), 中村友昭 (電通大), 萩原良信 (立命館), 小林一郎 (お茶大), 内田諭 (九大), 稲邑哲也 (NII), 井之上直也 (東北大), 岩橋直人 (岡山県立大)) によって今後のロボティクスに必要となる自然言語処理の展望をまとめたサーベイ論文 “Survey on frontiers on language and robotics” として, *Advanced Robotics* 誌に掲載されている [4]. 本稿でふれることのできない様々な話題については, この論文を参照されたい.

ニューラルネットが全盛の現在, 自然言語処理を用いる

原稿受付
キーワード：natural language processing, unsupervised learning, logical form, segmentation
*〒190-8562 東京都立川市緑町 10-3
*10-3 Midori-cho, Tachikawa City, Tokyo 190-8562

†https://robo-nlp.github.io/2017_index.html
††https://splu-robonlp.github.io/
†††https://www.aclweb.org/anthology/venues/robonlp/
‡http://www.emergent-symbol.systems/home

ロボティクスは End-to-End のニューラルネットを組むことで, 与えられた目的を最適化する研究が多い. しかし, こうしたアプローチには特有の限界もあると考えられる. そこで以下では, そうした現在の多くのアプローチとその問題点について述べ, 適切な知識表現によってそれらを解決する研究について紹介する. また, そうした方法の基礎となりうる筆者の共同研究について説明し, 今後の展望を議論する.

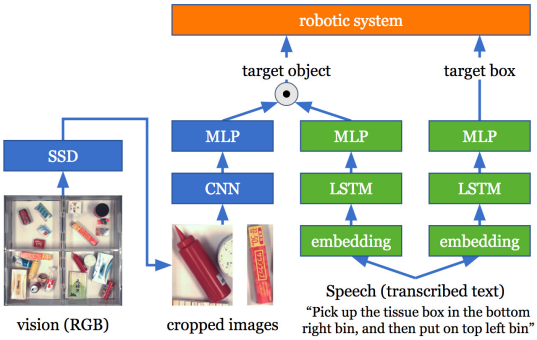
3. End-to-End 言語ロボティクスとその限界

言語ロボティクス研究で日本で最近最も話題になったのは, 2018 年の CEATEC に出展された, PFN 社の全自動片付けロボット†だと思われる. この基礎となる技術として, [5] では図 2(a) のように, 乱雑な箱の中から特定の物体を指示された場所に移動することを目的とし, ロボットが指示が曖昧な場合などに適宜人間に聞き返しながらタスクを遂行する. 学習アーキテクチャは図 2(b) のように完全にニューラルネットで組まれており, 指示発話を音声認識した結果は LSTM と後段のニューラルネットをブラックボックスとして最適化することで, 完全に End-to-End で動作することが特徴である. 論文では

> "move the rectangular object, with a green and white label, located in the middle of the top right box, to the top left box."

のように, かなり長い指示文†† でも適切に解析して, 正しい物体を目的の場所に移動できたことが報告されている.

(a) 箱の物体移動タスク. ここでは, 赤で囲まれた物体を右上の場所に移動することが自然言語で指示される.



(b) End-to-End のニューラルネットによるアーキテクチャ.

図 2 PFN 社の, 言語的指示による片付けロボット [5].

†https://projects.preferred.jp/tidying-up-robot/

†† こうした指示文は長いが, 代名詞や解釈が状況に依存する動詞などを含んでおらず, 曖昧性が非常に低い表現であることには注意する必要がある.

Go up the stairs and turn right. Go past the bathroom and stop next to the bed.

Walk all the way up the stairs, and immediately turn right. Pass the bathroom on the left, and enter the bedroom that is right there, and stop there.

図 3 VLN のデータセットにおける自然言語による指示 (一部) と, 初期位置からの画像の例. [7] による.

なお, 2019 年になって杉浦ら [6] は, 言語情報と画像情報に同時に張ったアテンションを駆使することで, [5] よりさらに高い精度を達成し, アテンションによってどの情報に注目しているかも分かる方法を提案した.

これらは決まった視点からの画像による研究であるが, PFN 社はその後上に述べた, 実際に空間を動き回って片付けを行うロボットの研究を行っている. また, Anderson らは Vision-and-Language Navigation (VLN) を 2018 年に提案し [7], 図 3 のようにシミュレータの仮想空間上で, 時々刻々得られる画像情報を用いて自然言語で与えられた経路を探索し, 実行するタスクを提案した. この場合の Room to Room(R2R) タスクにおいては, 自然言語で与えられた指示に従い, 初期位置から, ある部屋に移動することがタスクとなる. ロボットは指示文を LSTM に基づく seq2seq によって行動系列に書き換えつつ, 強化学習によって現実に近い環境の中で経路を探索する. このタスクの登場により, 自然言語処理の分野でも, 音声認識のように逆に発話を生成する確率を利用したり [8], 簡単なタスクから人間のように少しずつ学習したり [9] といった研究が進みつつある.

こうした End-to-End の方法は与えられたタスクにおいて高い精度を発揮できるが, 一方で多くの場合は指示文に対する正解の動作を必要とし, ドメイン外の場合には対応することができない. また, 明確な制約がある場合, あるいは眼前の画像を超えた論理的推論が必要な場合に対応できないという問題がある. たとえば, 家庭用ロボットにおいて「絶対に動かさないように運んで」「あの部屋は掃除しないで」といった指示は多くみられると考えられるが, こうした論理的な制約を, 言語に対して何の事前知識も持たない End-to-End なシステムが勝手に認識して必ず制約を守るように動作することは期待できない. このほか,

- 「その人」など代名詞を含む文
- 「きのう会った人に渡しておいて」のように埋め込み文を持つ関係節
- 「一番小さいコップに入れて持ってきて」といった論理操作

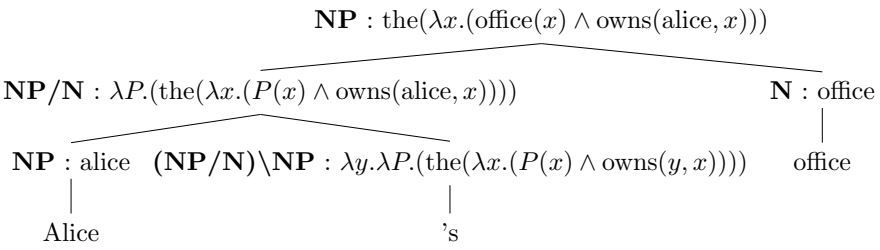


図 4 CCG による “Alice’s office” の構文解析と, 対応する λ 式. これにより, どちらも名詞である Alice と office が, 意味的にどのような関係にあるかが明示されている. ここでは NP と N を区別しているが, 一般には必ずしもその必要はない.

も, End-to-End なシステムが自動的に認識してくれるわけではない.

こうした意味で, **言語の構造的認識**はこれからのロボティクスには不可欠となると考えられる. こうしたアプローチはニューラルネットを作ることに比べて難しいため比較的数が少ないが, 様々な研究が行われており, 本稿ではそのうち構文および時間に関する 2 つの研究を紹介する.

4. 構造的言語処理とロボティクス

4.1 CCG による論理表現とロボット

Thomason ら [10] は, CCG (Combinatory Categorial Grammar, 組み合わせ範疇文法) によってロボットとの対話文を感覚情報とグラウンディングしつつ構文解析し, さらに対話によって知識を動的に更新してタスクを実行する方法を示した.

CCG [11] は文法のフォーマリズムの一つであり, CFG (句構造文法) のように, NP(名詞句) や VP(動詞句), PP(前置詞句) といった先験的なカテゴリを仮定しないのが特徴である. かわりにカテゴリとしては, 基本的に `S`(文) や `N`(名詞) といった自明なものしかなく, たとえば heavy のような形容詞は後ろに box のような名詞をとって全体で “heavy box” という名詞の働きをするため, `N/N` と表される. また, read のような他動詞は後に books のような名詞 `N` をとり, これによって次に前に主語となる名詞がくれば文 `S` になる動詞節 `S\N` となるため, `(S\N)/N` と表される.

このように,「何かの言葉を前後にとって何かになる」という操作は関数適用とみなすことができ, 図 4 のように, “’s” のような言葉に λ 式を結びつけておけば, 全体の意味表現は構文解析と同時に λ 式を関数適用する (β 簡約する) ことによって求められる. このようにして, CCG を用いれば, 構文解析と意味解析を自然に結びつけられるという点が最も大きな特徴である. 図 4 に, “Alice’s office” に対応する CCG の構文解析木と, その λ 式を示した. こうした解析により,

> Move the light mug from Bob’s office to the west, middle pod.

ような複雑な文も,

$$\begin{aligned}&\text{relocate}(\text{the}(\lambda x.(\text{lightweight}) \wedge \text{mug}(x))),\\ &\qquad\text{the}(\lambda y.(\text{office}(y) \wedge \text{owns}(\text{robert}, y))),\\ &\qquad\text{the}(\lambda z.(\text{west}(z) \wedge \text{middle}(z) \wedge \text{pod}(z))))\end{aligned}$$

のように解析することが可能になる [10]. 日本では, 戸次らによる `ccg2lambda` というシステム [12] が最もよく知られている.

もちろん, 実際にはどの単語がどんな λ 式を持つのかが全て分かっているわけではないから, 最小限の構文解析結果 ([10] では 44 個の指示文と λ 式のペア) から始め, 感覚情報を同時に用いて, 最も確率の高い構文解析結果と, λ 式による表現, および λ 式の変数がどの物体に対応しているかを自動的に学習し, 知識を増やしていく. 図 5 では, “the heavy mug” という表現がどのような λ 式で表されるのか (heavy を無視したり, mug が形容詞である可能性もある) を探索し, $\text{the}(\lambda x_i.(\text{and}(\text{heavy}(x), \text{mug}(x))))$ という λ 式の表現を得ている. なお, the とは, 条件を満たす唯一の実体を返す関数であり, ここではその実体が物体 o_1 である確率が 0.018, o_2 である確率が 0.982 と計算され, 結果としてこの λ 式全体が物体 o_2 と結び付けられている.

上記の実行にあたり, $\text{heavy}(o_1)$ などの述語が成り立つ確率は, 物体を操作した際の視覚, 触覚, 音情報からこれまでの認識結果を正例として SVM で識別器を学習し, 内部で使用している [13].

このように, 可能な物体の集合が与えられている, 述語の λ 式の一部が既知である (’s や the など) といった制限はあるものの, End-to-End のブラックボックスではなく, 確率的に不確実性も考慮しつつ, ロボットが内部で明示的に意味解析を行う研究も進みつつある. これにより, 固定された行動を事前に End-to-End で学んでおくだけでなく, ロボットが自ら何が不明であるのか人間に聞き, 自らの知識を動的に増やすことが可能になる. なお, こうした知能の枠組みは, 日本において岩橋・杉浦らが 2006 年に発表した LCore [14] [15] と類似しており, 本稿で説明した研究は, それを最先端の確率的な CCG や Active learning といった技術を用いて言い換えたものと考えることもできる.

$y \in \mathcal{P}(x)$	c_p	$g \in \mathcal{D}(\mathcal{P}(x))$	c_d	score(y)
$\text{the}(\lambda x_i.(\text{heavy}(x)))$	-1	o_1	0.364	$(g \neq o_2)$
		o_2	0.636	-1.45
$\text{the}(\lambda x_i.(\text{and}(\text{heavy}(x), \text{mug}(x))))$	-1.2	o_1	0.018	$(g \neq o_2)$
		o_2	0.982	**-1.22**
$\text{the}(\lambda x_i.(\text{mug}(x)))$	-1.2	o_1	0.111	$(g \neq o_2)$
		o_2	0.889	-1.32
$\text{and}(\text{heavy}, \text{mug})$	-1.2	$\text{and}(\text{heavy}, \text{mug})$	1	$(g \neq o_2)$

図 5 表現 “the heavy mug” に対応する λ 式の探索とスコア. ここでは 2 番目の λ 式が選ばれ, 学習データとして追加される.

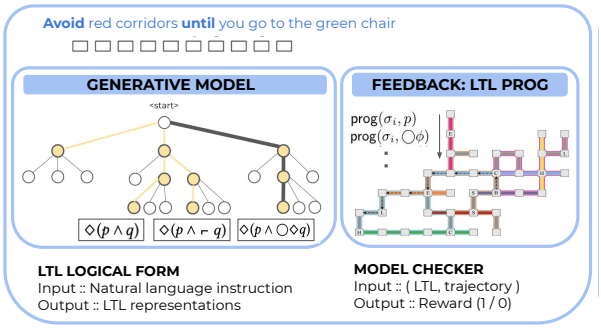


図 6 行動系列の背後に隠れた, LTL(線形時間論理) による論理式の探索と学習 [16].

4.2 LTL (線形時間論理) による制約

End-to-End な方法で必ずしも上手く扱えない現象として, 時間や順序に依存する制約がある. ロボットに対する指示として, 「台所を通ってからリビングに移動して」のように順序を指定する場合や, 「常に子供を避けて行動して」といった必ず守るべき制約を与える場合は多いと考えられる. 本稿冒頭で示した SpLU-RoboNLP 2019 でも発表された Patel らの研究 [16] は, 指示文に対応するこうした制約を LTL(線形時間論理) で表し, 指示文と行動のペアをデータとして, 論理式を潜在変数として学習する手法を提案した (図 6). これにより, ロボットが各状態で現在満たすべき制約を更新し, 必ず条件を満たす動作を行うことが可能になる. ニューラル手法では, このように「必ず条件を満たす」というような指定を自動的に行うことはできない.

LTL (Linear Temporal Logic) [17] は強化学習の文脈ですでに導入されている, 時間に関わる制約を表すことのできる論理であり, LTL の論理式 ϕ は次の形式で定義される.

$$\phi ::= \pi \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \diamond\phi \mid \Box\phi \mid \bigcirc\phi \mid \phi \cup \psi$$

ここで $\neg, \wedge, \vee$ は通常通りの not, or, and であるが, その他に $\diamond$ は eventually を, $\Box$ は globally を, $\cup$ は until を, $\bigcirc$ は next をそれぞれ意味する. よって, `is_corridor` のような場所を表す述語を p, q とすると, たとえば $(\diamond p) \wedge (\bigcirc \diamond q)$ は「最終的に p にいて, 次に最終的に q にいる」という条件を表す.

このとき, ロボットが p を通過すると, この条件は $\diamond q$ に「進める」ことができるため, ある論理式に従うはずの多数の行動履歴を集めれば, そこからその論理式を逆算することができる. 具体的には, 自然言語による指示を seq2seq で後置記法で表した LTL の論理式に変換し, その論理式は強化学習を通じて, 各時刻で行動と合っていれば 1, 合っていなければ 0 の報酬を受け取ることで学習される. このとき, 常に〜を満たせ ($\Box$) という制約は, すべての時点で論理式が成り立つことを要求するとして処理することができる.

2 節でみたように, 言語による指示を seq2seq で直接, 運動系列に変換することも可能ではあるが, こうして論理表現を介することで解釈性も上がり, 様々な制約を人間が明示的にロボットの動作に取り入れることが可能になる. 実験により, 時間的順序を含む命令について, この方法は単純な seq2seq より高精度であることが確認されている. また, 九日付の英有力紙タイムズは、同国南部のウェイマスに近いポートランドの海軍基地を欧州向け物資の陸揚げ基地として日本企業ないし企業連合にそっくり売却する構想が浮上していると報じた。五輪五位の清水宏保はインカレも 2 種目を制しており、堀井にどこまで迫るか。

図 8 京大コーパスの教師なし形態素解析の結果の例. 辞書や教師データはまったく使用していない.

図 7 のように CCG から得られた λ 式による論理表現より, 順序に関する制約をコンパクトに表すことができる.

```
LTL: at_lamp U ◇ (at_chair)

CCG: λa.move(a) ∧ post(a,intersect
(λx.chair(x),you))∧pre(a,front(you,
λx.lamp(x)))
```

図 7 「ランプがある椅子の前まで移動する」を表す LTL と CCG の論理表現. LTL の方が, 「〜まで」といった順序を含む条件をコンパクトに表すことができる.

5. ロボティクスにおける動作の分節化

5.1 形態素解析と運動データ

こうした中で筆者は, ロボットの連続的な動きを「動作」に分節化することを中心として, 電通大・阪大・お茶大のグループと共同研究を進めてきた. 図 9 に示したように, 提案法ではロボットの運動データを, 完全に教師なしで「動作」の系列に分解することができ, ロボティクスだけでなく, 舞踊学, 動物学 [18] などにも有用な基礎技術と考えられる.

この問題は, 日本語や中国語のような言語において “ロボットは歩いた” のような文字列を「ロボット/は/歩い/た」と単語に分割する (+場合により, 品詞等の情報を付与する) **形態素解析**といわれるタスクと同じであり, 共同研究は言ってみれば, ロボットの運動の「形態素解析」をしているといえる. 通常はこの問題は, 言語の場合は上記のような文字列→単語分割結果の正解データを大量に準備し, そこから SVM や CRF, 最近ではニューラルネットなどを使って分割を教師あり学習することで解くことができる.

しかし, ロボットや人間の動作についてそうした「正解」を用意するのは, 基準が明確でない上に, 言語のようにかなり厳密な規則性があるとはいえず†, 困難である. そのため, 何らかの方法で, 教師なしで自動的に「動作」を学習する必要があると考えられる.

実は言語においてもこの問題は重要であり, 未知の言語や新語を自動的に認識できる教師なしの単語分割は, 長く未解決問題であった.†† このような中, 筆者は 2009 年に, ノンパラメトリックベイズ法と動的計画法, MCMC を組み合わせることで文を「単語」に自動的に分解する教師なし形態素解析を発表した [19]. これは文字–単語の階層マルコフ言語モデルであり, 文字の ∞-グラムモデルからまず「単

†自然言語では, 「彼をの言っつ」「側アる聞のと」のように非文法的な文は許容されず, ほとんど出現しないという強い制約がある.
††これは重要な問題であるため, それまでにもいくつかの試みがあったが, ヒューリスティックを用いたり, EM アルゴリズムのような最尤推定を基にしていたため, 局所解に陥ってしまい, 意味のある結果が得られていなかった.

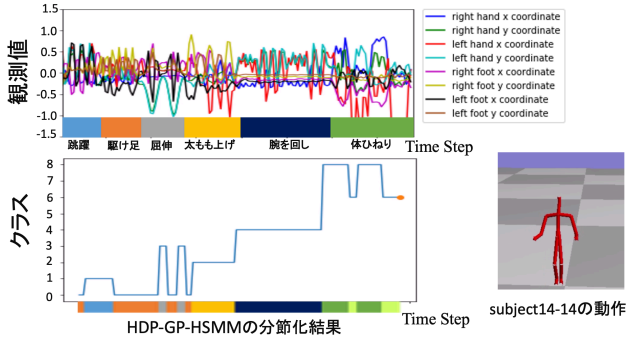


図 9 HDP-GP-HSMM による運動の分節化．上段の各関節の運動軌跡が，統計モデルにより自動的に，下段のように「動作」の系列に分解されている．

1 神戸では異人館 街の 二十棟 が破損した 。
2 神戸 では 異人館 街の 二十棟 が破損した 。
10 神戸 では 異人館 街の 二十棟 が破損した 。
50 神戸 で は異人 館 街 の 二 十 棟 が 破損 し た 。
100 神戸 で は 異 人館 街 の 二 十 棟 が 破損 し た 。
200 神戸 で は 異人館 街 の 二 十 棟 が 破損 し た 。

図 10 教師なし形態素解析の学習過程．MCMC 法に基づき，観測された文字列の単語分割を繰り返し確率的にサンプリングすることで言語モデルを改善し，学習を行う．

語」の綴りが無限個生成され，それを n グラムモデルで組み合わせることで文が生成されたと考え，文の文字列から「単語」を逆に推論する．学習の過程は，図 10 のようになる．まず，ランダムな単語分割あるいは「文」が一つの単語となる自明な分割から始め，上記の言語モデルを学習する．次に学習データの各文に対して新しい単語分割を動的計画法を用いて確率的にサンプリングし，それによって言語モデルを更新する．以上をモデルが収束するまで繰り返す．これにより，例えば京大コーパス (毎日新聞の記事の一部) の文字列のみから，図 8 のように「単語」への分割を辞書や教師データを使わず，教師なしで学習することができる．

5.2 単語から動作へ

この方法を応用すれば，ロボットの運動も「動作」(＝単語) に分解することができる．最初に離散化して言語と同様に記号列にすることも可能であるが [20]，我々はここでガウス過程 [21] を用い，単語の綴りにあたる各「動作」の軌道が図 11 の Z_j のように，状態ごとのガウス過程から生成されたと考える．運動データは図 11 の S のような軌道をなすから，この問題は全体の軌道を，言語と同様に「単語」，すなわち動作に分解する問題であり，図 10 と同様な方法で，MCMC 法により分割を繰り返し推定する．

統計的には，こうした分割問題は隠れセミマルコフモデル [22] と呼ばれ，通常の隠れマルコフモデルと異なり，状態が 1 つ前の時刻の観測値だけに依存しないのが特徴である．[1] ではさらに，階層ディリクレ過程 [23] を用いて潜在状態の数も自動的に推定する HDP-GP-HSMM を提案し，適切な状態数が推定されることを確かめている．

5.3 VAE による潜在空間での分割

上記の研究では，観測値としては肩や肘といった代表的な数次元の関節角の時系列を用いていたが，全身の動作を表現するためには，高次元の運動をより低次元に縮約して

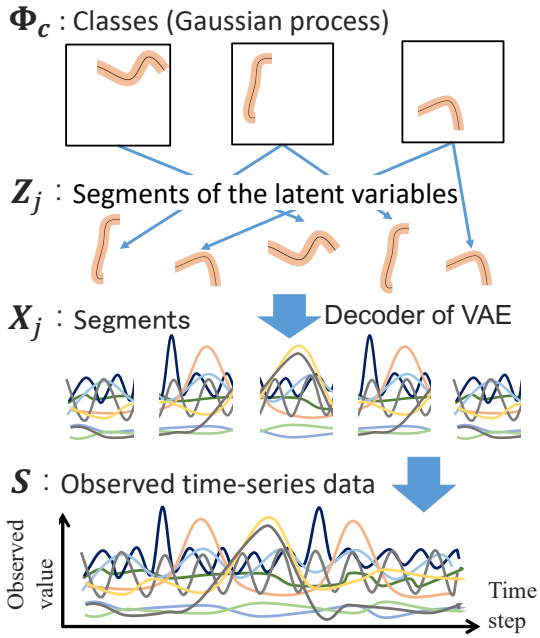


図 11 HVGH による，VAE で圧縮された潜在空間での軌跡のガウス過程による分節化．

表現する必要がある．線形モデルである主成分分析はこの目的には適しておらず，確率的主成分分析のカーネル拡張であるガウス過程潜在変数モデル (GPLVM) [24] を使うと，高次元の運動をより本質的な次元に圧縮できることが知られている [25]．我々はこの考察をもとに，GPLVM とほぼ等価なニューラルネットによって図 11 のように観測時系列を低次元の潜在空間に圧縮し，この潜在空間の時系列を分節化する HVGH† [2] をさらに提案した．HVGH では潜在系列が分節化された後に，潜在空間での状態ごとのガウス過程だけでなく，観測時系列を低次元に圧縮するニューラルネットも VAE を用いて最適化し，これらの学習を交互に行う．実験により，図 12 のように 96 次元の関節角データをフルに利用して学習を行うことができ，VAE によって圧縮された潜在空間において，それぞれの動作は一定の軌跡を通ることを明らかにした．

6. 今 後 の 課 題

End-to-End アプローチを補完する，自然言語の構造的な

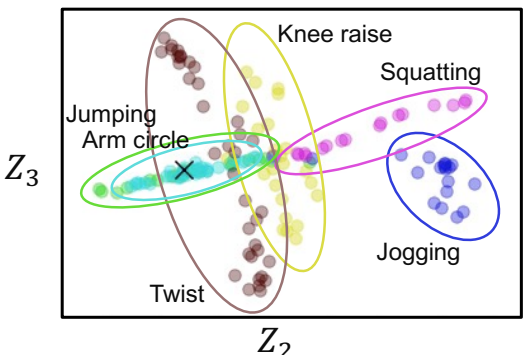


図 12 HVGH による潜在空間で分節化された動作の軌跡．各フレームに対応する点は時系列で連続している．

†Hierarchical Dirichlet process–Variational autoencoder–Gaussian process–Hidden semi-Markov model の略．

認識を使用したロボティクスについて紹介した. 構造的なアプローチではあるものの, 従来そのために用いられていた古典的な教師あり学習とは異なり, これらは構文木や論理式, 単語分割といった人手によるアノテーションを必要とせず, **ロボットが自ら学習する**点が特徴である.

一方で, ロボットに必要な構造的な認識はそれら以外にも, 3 節で述べたように代名詞や関係節を含む文, 仮想的な条件を含む文 (「もし雨が降りそうなら, 洗濯物を取り入れておいて」), メタファーの理解など, 多数の側面があると考えられる. 研究はこれまで, それぞれの側面に注目して進められてきたが, 実際にはそれらのモジュールを統合する学習が必要になり, また開発もモジュール別に可能になることが望まれる. これを可能とする中村・谷口らの Serket [26], Neuro Serket [27] はこの意味で, 非常に重要なフレームワークだと考えられる. VLN のように現実のロボットの動作に近いシミュレーション環境とタスクも整ってきている現在, より多くの自然言語処理研究者がロボティクスに関わることで, 問題を統合的に解決し, 言語の中だけでは解けない言語理解が可能になることを筆者は期待している.

謝 辞 共同研究者の皆様と LangRobo 研究会のメンバー, および研究会に参加された皆様に感謝いたします. 本研究は, 日本学術振興会科研費 26280096 および 18H03295 の支援を受けて行いました.

参 考 文 献

[1] Masatoshi Nagano, Tomoaki Nakamura, Takayuki Nagai, Daichi Mochihashi, Ichiro Kobayashi, and Masahide Kaneko. Sequence Pattern Extraction by Segmenting Time Series Data Using GP-HSMM with Hierarchical Dirichlet Process. In *IROS 2018*, pages 4067–4074, 2018.

[2] Masatoshi Nagano, Tomoaki Nakamura, Takayuki Nagai, Daichi Mochihashi, Ichiro Kobayashi, and Wataru Takano. High-dimensional Motion Segmentation by Variational Autoencoder and Gaussian Processes. In *IROS 2019*, pages 105–111, 2019.

[3] Daichi Mochihashi. Gaussian Process Generative Models for Language and Robotics, 2019. CoRL 2019 Tutorial.

[4] T. Taniguchi, D. Mochihashi, T. Nagai, S. Uchida, N. Inoue, I. Kobayashi, T. Nakamura, Y. Hagiwara, N. Iwahashi, and T. Inamura. Survey on frontiers on language and robotics. *Advanced Robotics*, 33(15–16):700–730, 2019.

[5] J. Hatori, Y. Kikuchi, S. Kobayashi, K. Takahashi, Y. Tsuboi, Y. Unno, W. Ko, and J Tan. Interactively Picking Real-World Objects with Unconstrained Spoken Language Instructions. In *ICRA 2018*, pages 3774–3781, 2018.

[6] A. Magassouba, K. Sugiura, and H. Kawai. A Multimodal Target-Source Classifier with Attention Branches to Understand Ambiguous Instructions for Fetching Daily Objects. *IEEE Robotics and Automation Letters*, 5:532–539, 2020.

[7] Peter Anderson, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sünderhauf, Ian Reid, Stephen Gould, and Anton van den Hengel. Vision-and-Language Navigation: Interpreting Visually-Grounded Navigation Instructions in Real Environments. In *CVPR 2018*, pages 3674–3683, 2018.

[8] Shuhei Kurita and Kyunghyun Cho. Generative Language-Grounded Policy in Vision-and-Language Navigation with Bayes' Rule. In *ICLR 2021*, page to appear, 2021.

[9] Wang Zhu, Hexiang Hu, Jiacheng Chen, Zhiwei Deng, Vihan Jain, Eugene Ie, and Fei Sha. BabyWalk: Going Farther in Vision-and-Language Navigation by Taking Baby Steps. In *ACL 2020*, pages 2539–2556, 2020.

[10] Jesse Thomason, Aishwarya Padmakumar, Jivko Sinapov, Nick Walker, Yuqian Jiang, Harel Yedidsion, Justin Hart, Peter Stone, and Raymond J. Mooney. Jointly Improving Parsing and Perception for Natural Language Commands through Human-Robot Dialog. *The Journal of Artificial Intelligence Research (JAIR)*, 67:327–374, 2020.

[11] Mark Steedman. *The Syntactic Process*. Language, Speech, and Communication. MIT Press, 2000.

[12] Pascual Martínez-Gómez, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki. ccg2lambda: A Compositional Semantics System. In *ACL-2016 System Demonstrations*, pages 85–90, 2016.

[13] Saeid Amiri, Suhua Wei, Shiqi Zhang, Jivko Sinapov, Jesse Thomason, and Peter Stone. Multi-modal Predicate Identification using Dynamically Learned Robot Controllers. In *IJCAI-18*, 2018.

[14] Naoto Iwahashi. Robots That Learn Language: A Developmental Approach to Situated Human-Robot Conversations. In *International Workshop on Emergence and Evolution of Linguistic Communication*, pages 143–167, 2006.

[15] 杉浦孔明, 岩橋直人, 柏岡秀紀, 中村哲. 言語獲得ロボットによる発話理解確率の推定に基づく物体操作対話. 日本ロボット学会誌, 28(8):978–988, 2010.

[16] Roma Patel, Stefanie Tellex, and Ellie Pavlick. Learning to Ground Language to Temporal Logical Form. In *SpLU-RoboNLP 2019*, 2019.

[17] Zohar Manna and Amir Pnueli. *The Temporal Logic of Reactive and Concurrent Systems: Specification*. Springer, 1992.

[18] 三村喬生, 中村友昭, 松本惇平, 西条寿夫, 須原哲也, 持橋大地, 南本敬史. 霊長類における身体動作時系列の分節推移構造推定. In 2019 年度人工知能学会全国大会, pages 1C4–J–3–01, 2019.

[19] Daichi Mochihashi, Takeshi Yamada, and Naonori Ueda. Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling. In *Proceedings of ACL-IJCNLP 2009*, pages 100–108, 2009.

[20] Tadahiro Taniguchi, Shogo Nagasaka, Kentarou Hitomi, Naiwala P. Chandrasiri, Takashi Bando, and Kazuhito Takenaka. Sequence Prediction of Driving Behaviour Using Double Articulation Analyzer. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(9):1300–1313, 2015.

[21] Carl Edward Rasmussen and Christopher K. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[22] Kevin Murphy. Hidden semi-Markov models (segment models), 2002. http://www.cs.ubc.ca/˜murphyk/Papers/segment.pdf.

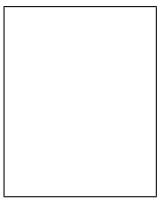
[23] Yee Whye Teh, Michael I. Jordan, Matthew J. Beal, and David M. Blei. Hierarchical Dirichlet Processes. *JASA*, 101(476):1566–1581, 2006.

[24] Neil Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *Advances in Neural Information Processing Systems*, pages 329–336, 2004.

[25] K Grochow, S.L. Martin, A. Hertzmann, and Z. Popović. Style-based Inverse Kinematics. *ACM transactions on Graphics*, 23(3):522–531, 2004.

[26] Tomoaki Nakamura, Takayuki Nagai, and Tadahiro Taniguchi. SERKET: An Architecture For Connecting Stochastic Models to Realize a Large-Scale Cognitive Model. *Frontiers in Neurorobotics*, 12(25), 2018.

[27] Tadahiro Taniguchi, Tomoaki Nakamura, Masahiro Suzuki, Ryo Kuniyasu, Kaede Hayashi, Akira Taniguchi, Takato Horii, and Takayuki Nagai. Neuro-SERKET: Development of Integrative Cognitive System through the Composition of Deep Probabilistic Generative Models. *New Generation Computing*, 84, 2019.



持橋 大地（Daichi Mochihashi）

統計数理研究所 数理・推論研究系 准教授. 2005 年奈良先端大・情報・博士後期課程修了. 博士 (理学). ATR 音声言語コミュニケーション研究所, NTT コミュニケーション科学基礎研究所各研究員を経て現職. 統計的自然言語処理と機械学習, 特に意味的処理と連続系との接続に興味を持つ.