柔軟多脚型ロボット : TAOYAKA-S II : ソフトロボットの物理的性質を用いた強化学習における状態・行動空間の抽象化

本間, 義大 / HOMMA, Yoshihiro

(出版者 / Publisher)
法政大学大学院理工学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 理工学・工学研究科編
(巻 / Volume)
60
(開始ページ / Start Page)
1
(終了ページ / End Page)
7
(発行年 / Year)
2019-03-31
(URL)
https://doi.org/10.15002/00022030

柔軟多脚型ロボット: TAOYAKA-S Ⅱ -ソフトロボットの物理的性質を用いた 強化学習における状態・行動空間の抽象化-

THE SOFT MULTI-LEGGED ROBOT TAOYAKA-S II -ABSTRACTION OF THE STATE-ACTION SPACE OF REINFORCEMENT LEARNING USING THE PHYSICAL PROPERTIES OF A SOFT ROBOT-

本間 義大
Yoshihiro HOMMA
指導教員　伊藤 一之 教授

法政大学大学院理工学研究科電気電子工学専攻修士課程

This paper considers the abstraction of the state-action space of reinforcement learning using the physical properties of a soft body. In general, soft robots can adapt to complex environments owing to their flexibility. This adaptability is utilized for abstracting the state-action space. The policy acquired using the abstraction was found to have generality, and to greatly reduce the size of the state-action space. The proposed framework was applied to the soft multi-legged robot TAOYAKA-S II, and demonstrated that the robot could easily acquire an effective policy moving within a given environment. Experiments were conducted to demonstrate climbing motion over a pipe and walking motion over a flat surface. The proposed framework made the policy applicable to other columnar objects without requiring additional learning.

***Key Words** : reinforcement learning, state-action space, flexibility, multi-legged,*

１．はじめに

近年, 強化学習, 及びソフトロボットの両方が大きな注目を集めている. 強化学習は, 教師を必要とせず, 単独で学習できるため, ロボット学習に向いている. 加えて, 強化学習を使用することで, 適応性の高いコントローラが実現可能となる. また, ソフトロボットは, 自身の高い柔軟性を利用することで, 環境にうまく適応することが出来る. 強化学習とソフトロボットを組み合わせることで, 適応性の高い自律型ロボットを実現することが可能となる.

しかし, 一般に, 自由度が大きいため, 強化学習をソフトロボットに適用することは難しい. また, 状態・行動空間, 及び学習完了までの時間は, 自由度の増加とともに指数関数的に増加する. さらに, 通常の強化学習アルゴリズムは一般化能力を持たないため, たとえ有用な方策が得られたとしても, その方策は新しい環境には適用できない.

したがって, 学習時間の短縮, 及び一般化を達成するための状態・行動空間の抽象化は, ソフトロボットに強化学習を適用するための最も重要な課題であると考えられる.

従来研究では, 状態・行動空間を抽象化するための様々なアプローチが提案されてきた [1-7]. しかし, これらのアプローチは, 抽象化のために多大な計算コストを要するため, 多くの自由度を有するロボットに強化学習を適用することは困難であった.

一方, タコやヘビなどの自然界の生き物は, 高い自由度を有するにもかかわらず, 比較的小さな脳を使って実際の複雑な環境で巧みに学び行動することができる. 抽象化と一般化のために生物が使用する学習メカニズムは, 未だに解明されていない. しかし, 生物の柔軟な身体が重要な役割を果たすと考えられている [8-11].

本研究では, 柔軟な身体を利用して状態・行動空間を抽象化し, 提案する枠組みを柔軟多脚型ロボット (TAOYAKA-S II)に適用する. ロボットが与えられた環境内で移動するための効果的な方策を容易に得られることを確認する. パイプと水平面の 2 つの異なる環境下で学習により方策を獲得し, また, パイプを用いた学習で得られた方策が, 追加学習なしに, 四角柱などの他の柱状物体においても適用可能であることを実験により確認することで, 一般性があることを示す.

２． 提案する枠組み

提案する枠組みを Fig.1 に示す [12]．この枠組みでは，状態・行動空間は柔軟な身体のダイナミクスを用いて抽象化されている．学習モジュールで使用される状態・行動空間のサイズは非常に小さいため，ロボットは妥当な制限時間内に学習を完了することが可能となる．また，環境と柔軟な身体との相互作用に応じてロボットの行動が生成されるため，複雑な環境でも追加学習なしにロボットが適応的に行動することが出来る．ロボットの詳細は第 3 項に示す．

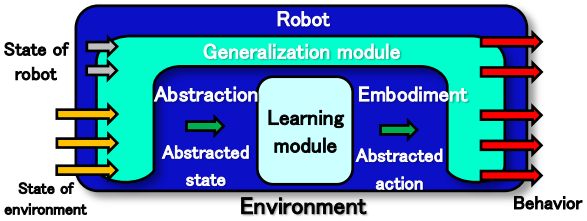


Fig.1 Proposed framework

３． 柔軟多脚型ロボット：TAOYAKA-SII

開発したロボット: TAOYAKA-S II を Fig.2 に示す．ロボットは，柔軟な身体と環境との間の相互作用を利用して，環境に適応するための多くの複雑な行動からいくつかの本質的な行動を抽出するように設計されている．

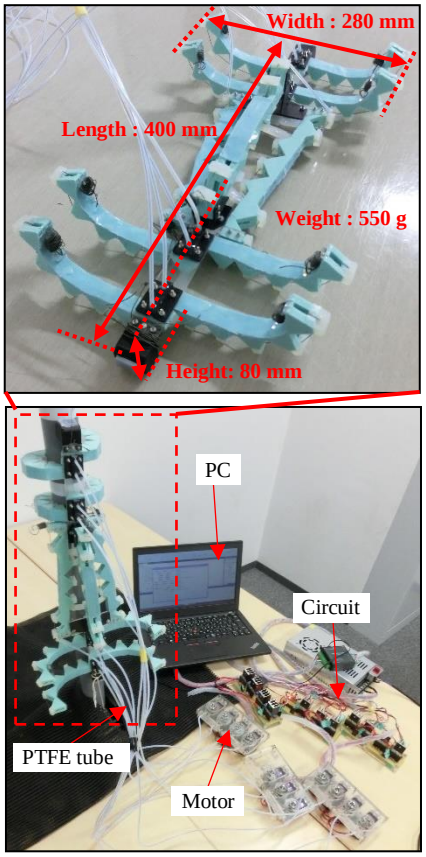


Fig.2 TAOYAKA-S II

ロボットは，8 本の柔軟脚と 3 本の柔軟リンクで作製された体幹で構成されている．脚と体幹を収縮させるために，プラスチック製の紐が，PTFE チューブを介して，モーターに接続されている．モーターは PC によって制御され，モーターに取り付けられたプーリーが紐を引くことによってロボットが動作する．

（１）柔軟脚

TAOYAKA-S II に搭載された柔軟脚を Fig.3 に示す．硬さ 60 A のシリコンで作製されており，6 つの節を有する．また，内部には，脚を閉じるためのプラスチック製の紐が，PTFE チューブを介して取り付けられている．対象物との間に十分な摩擦を実現するために，柔らかいシリコンブロック (30 A)が脚の内側に取り付けられている．また，紐を引いた際に，紐を引く力が，柔軟脚自身の伸びる力に変換されてしまうことを防ぐために，脚の外側にリボンが取り付けられている．

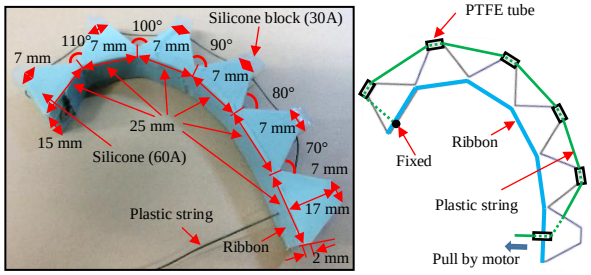


Fig.3 TAOYAKA-SII’s flexible leg

各節間の角度は，Fig.3 に示すように，根元から先端に向けて徐々に大きくなるように設定されている．これにより，関節を閉じるために必要な力は，根元から先端に向けて大きくなる．この物理的性質のため，脚は根元から徐々に屈曲する (Fig.4)．この振る舞いは，タコの把持戦略を模倣したものである．タコは物体を掴む際，触手を根元から先端に向けて徐々に触れさせていることが報告されている．この戦略により，タコは常に対象物の形状を感知することなく，未知の物体を掴む，もしくは包むことができる [13-14]．この動作を単純なメカニズムで実現した．

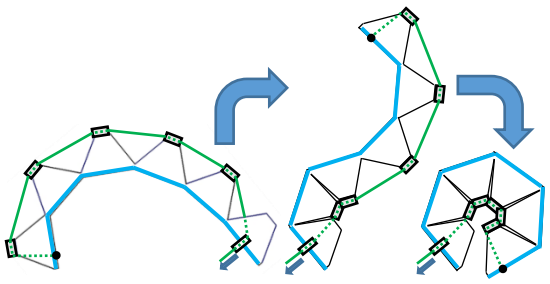


Fig.4 Movement of the flexible leg

（２）体幹

体幹を Fig.5 に示す．体幹は 3 つの柔軟リンクで構成されており，リンクを収縮させることで，ロボットは三次元方向への収縮運動を実現することが可能となる．体幹は，リンクが機体を支える十分な力を確保するために，硬さ 60 A のシリコン，及びリボンで作製されている．体幹を収縮させるために，各リンクには，Fig.6 に示すように，柔軟脚と同様な紐のメカニズムが採用されている．

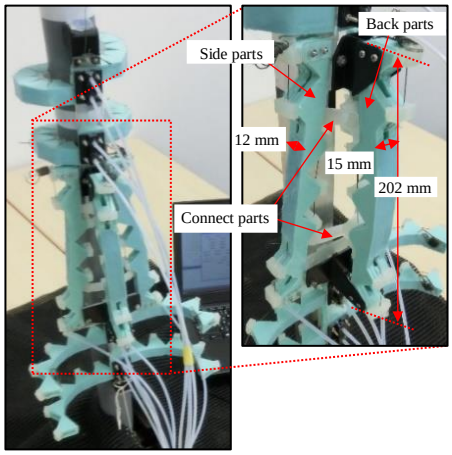


Fig.5 Developed trunk

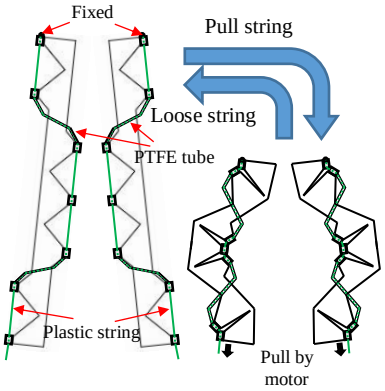


Fig.5 Movement of the trunk

（３）状態・行動空間の抽象化

身体のメカニズムを使うことで，状態・行動空間は，紐の状態と動きを使用して記述することが可能となる．例えば，脚に取り付けられた紐を引くことで，脚は閉じ，その形状は自律的に環境に適応する．これはソフトロボットの最も重要な利点である．言い換えると，脚の多自由度を制御するために必要な大きな処理は，力学的特性によってリアルタイムで行われる．したがって，学習モジュールは，多くの自由度を制御する必要がなく，単に紐を引いたり緩めたりするという動作だけで済み，行動空間の大きさを大幅に縮小することが出来る．同様に，学習モジュールは，紐の状態を利用することで抽象化された状態を観測することが出来る．脚に取り付けられた紐を引くだけで，脚は対象物を把持できるため，学習モジュールは，対象物の形状を知る必要はない．したがって，状態空間の大きさをかなり小さくすることが出来る．

状態空間の大きさが縮小されるだけでなく，抽象化された状態と行動を使用して得られた方策にも一般性がある．学習モジュールが対象物を把持するための方策を学習する場合，物体の形状や大きさは，柔軟な身体によって吸収されるため，その方策は追加学習なしで他の対象物に適用することが可能となる．

状態・行動空間の構成に関する詳細な説明は第 5 項に示す．

４． 学習モジュール

柔軟な身体によって実現される汎化能力を確認するために，式(1)のような典型的な Q 学習を採用する [15]．s は状態，a は行動で，a によって s から遷移した状態が s’．s’の時に選択する行動が a’である．$\max_{a'} Q(s',a')$は，遷移した状態で選択できる行動に対する Q 値の中で最大の値である．また，α は学習率，r は報酬，γ は割引率である．このアルゴリズムは非常に単純であるが，一般化能力は備えていない．

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + \alpha\left\{r(s,a) + \gamma \max_{a'} Q(s',a')\right\} \quad (1)$$

５． 実ロボットを使用した学習

本研究では，TAOYAKA-S II を用いて，柱状物，及び水平面という 2 つの環境下に置かれた場合の学習を行った．初めに，1 つのパイプを用いて学習した．また，このとき得られた方策を使用し，追加学習なしで他の柱状物に適用可能か検証した．次に，Q 値を初期化し，柱状物を用いた学習に使用したものと同じアルゴリズムを用いて，水平面上で学習を行った．

ロボットの状態を Table 1，動作を Table 2 にそれぞれ示す．

Table 1 State

State of the robot S_{ijk}

k: State of the upper leg

No.	0	1	2
state	closed	neutral	opened

i: State of the upper leg

No.	0	1	2
state	closed	neutral	opened

j: State of the trunk

No.	0	1
state	shrinked	stretched

S	Upper leg			Trunk	
	closed	neutral	opened	shrinked	stretched
No.	0	1	2	0	1
Fig.					

S	Lower leg		
	closed	neutral	opened
No.	0	1	2
Fig.			

Table 2 Action

No.	action
0	pull the string of the upper leg to the limit
1	pull the string of the upper leg to the neutral position
2	loosen the string of the upper leg
3	pull the string of the trunk
4	loosen the string of the trunk
5	pull the string of the lower leg to the limit
6	pull the string of the lower leg to the neutral position
7	loosen the string of the lower leg

本来であれば，このソフトロボットは自由度が高く，状態・行動空間は非常に大きい．しかし，第 2，3 項で述べたように，このロボットはその柔らかい身体の力学的特性を利用して様々な環境に適応する．したがって，学習モジュールは身体の正確な状態を知る必要はない．同様に，ロボットの振る舞いは，紐を引く，緩めるという単純な動作でロボットを動かすだけで，ソフトロボットと環境との間の相互作用によって生成される．したがって，学習モジュールはロボットの多自由度を正確に制御する必要はない．

このように，状態・行動空間をコンパクトに構成することが出来る．言い換えれば，柔軟な身体は，正確で大きな状態・行動空間を小さなものへと抽象化しているといえる．

Table 1 に示すように，脚の状態は 3 つと設定する．この表中で，脚が閉じている状態は，紐が限界まで引っ張られている状態で．開いている状態は，紐が緩み，柔軟脚が初期の形状になっている状態を意味する．中間は，開いている状態と閉じている状態の間である．同様に，体幹は，Table 1 に示すように，2 つの状態が考えられる．ロボットの状態はこれらの組み合わせであり，状態数は 18 である．このサイズは，従来の方法のものよりも著しく小さい．また、行動は，Table 2 に示すように，8 つの紐の操作を設定する．以上より，状態・行動空間は 144 となり，非常に小さなサイズとなっていることがわかる．

実験環境を Fig.7 に示す．移動距離を測定するために光学式モーションセンサーを使用した．センサーはワイヤーを介してロボットに接続され，ロボットが前進するとワイヤーがセンサーを引き，後退すると，おもりがセンサーを反対方向に引っ張る．移動距離が正の場合は報酬として 30 を与え，負の場合は-30 の報酬を与えた．また，移動距離がほぼゼロの場合の報酬を-1 とした．

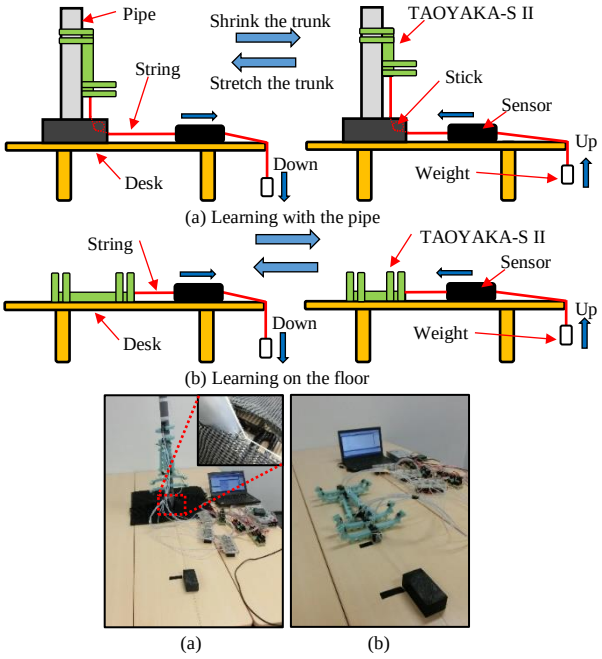


Fig.7 The experiment environment

６． 学習結果

本実験では，学習率を 0.5，割引率を 0.9 に設定した．行動の選択方法には，ε-greedy 法を用いた．行動をランダムに選択する確率を 10 %とし，1 試行は，行動を 20 回選択するまでとし，学習曲線が収束するまで試行を繰り返した．

（１）円柱状物体の登攀

直径 49 mm のパイプを用いての学習を行った．学習曲線を Fig.8 に，初期の状態，1 回目，10 回目，20 回目の試行終了時の位置を Fig.9 に示す．また，得られた方策を Fig.10，状態遷移図を Fig.11 に示す．学習により移動距離が増加することを確認し，最終的には適切な登攀パターンが得られたことを確認した．また，学習に必要な試行回数は 20 回であった．これは，状態・行動空間の抽象化が学習時間の短縮に非常に効果的であることを意味していると考えられる．

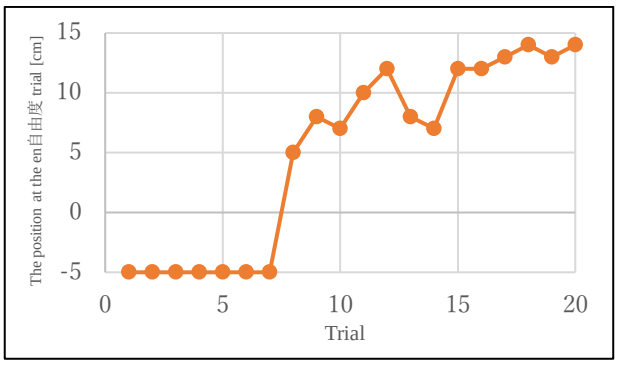


Fig.8 Learning curve with the pipe

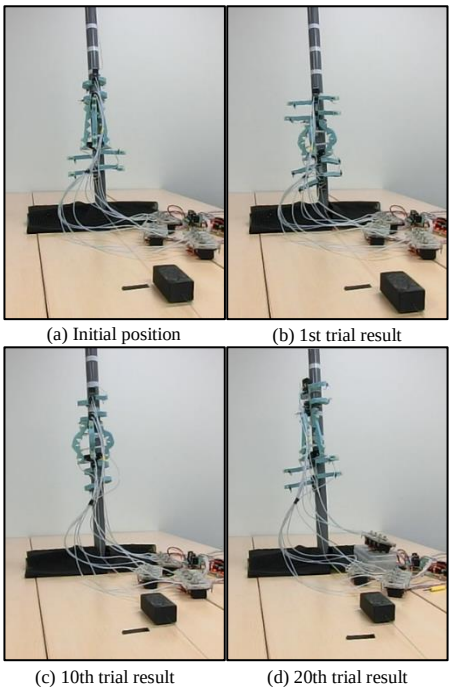


Fig.9 Moving distance at each trial (climbing)

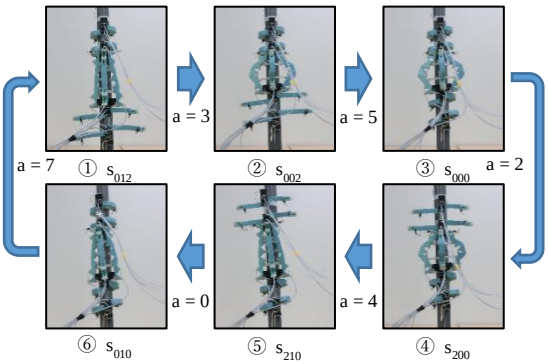


Fig.10 The obtained climbing pattern

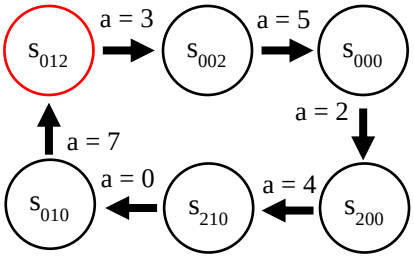


Fig.11 The state transition diagram (climbing)

次に，得られた登攀パターンを他の種類の柱状物に適用して汎化能力が備わっているか確認した．実験には，学習時に使用したものよりも細いパイプ，四角柱，二本のパイプの組み合わせ，天然の木を使用した．実験結果を Fig.12-15 に示す．この実験より，学習により得られた方策は，追加学習なしに様々な柱状物に適用可能であることがわかる．提案した枠組みにより，一般性を有することを確認した．

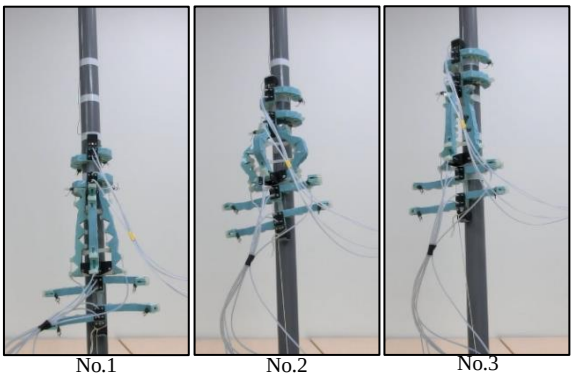


Fig.12 Experiment result (Pipe: diameter of 44 [mm])

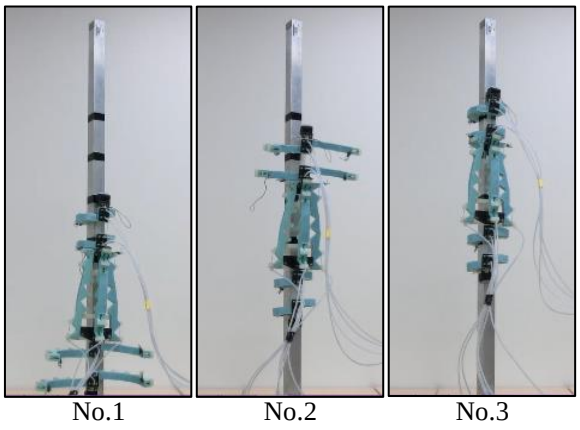


Fig.13 Experiment result (Square pillar: 30 × 30 [mm])

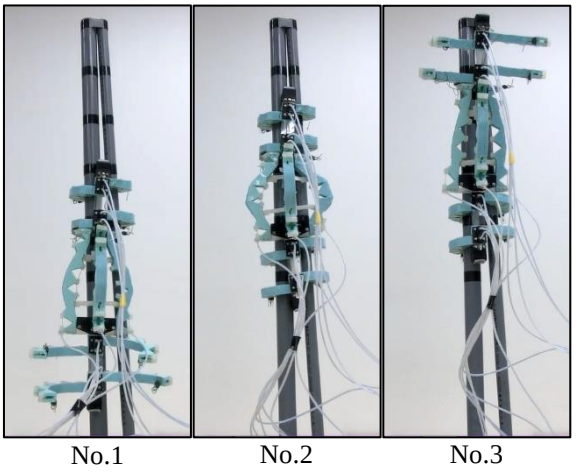


Fig.14 Experiment result (Two pipes: d = 34 [mm] & 22[mm])

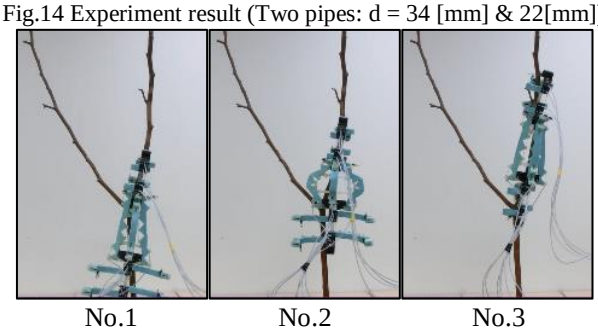


Fig.15 Experiment result (Nature tree)

（２）水平面の歩行

提案した枠組みの汎用性を確認するため，異なる環境の例として，水平面においての学習を行った．このタスクでは，ロボットが元の状態に復帰できない場合は失敗として，手動で負の報酬を与えることとする (Fig.16)．その他の設定はパイプにて学習した時に使用したアルゴリズムと同じものを使用した．

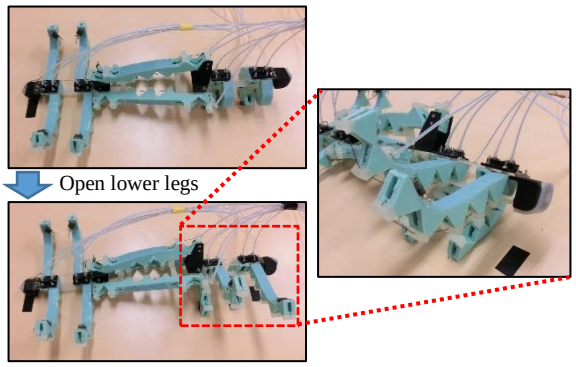
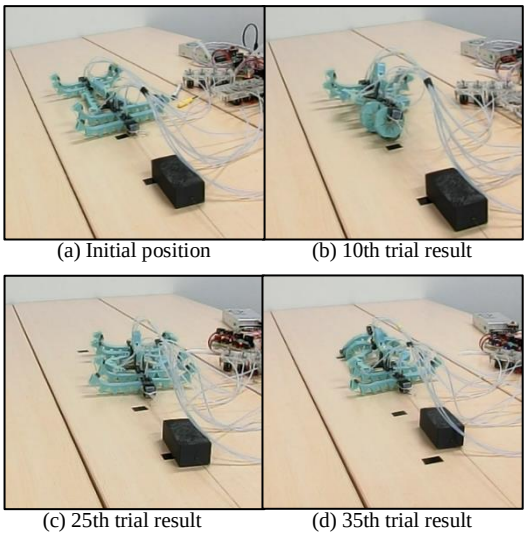


Fig.16 Failure pattern

初期の状態，10, 25, 35回目の試行終了時の位置を Fig.17，学習曲線を Fig.18 に示す．また，得られた方策を Fig.19，状態遷移図を Fig.20，実現した歩行行動を Fig.21 に示す．



(a) Initial position (b) 10th trial result

(c) 25th trial result (d) 35th trial result

Fig.17 Moving distance at each trial (walking)

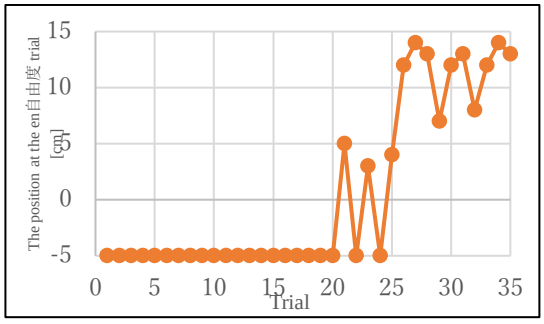


Fig.18 Learning curve on the floor

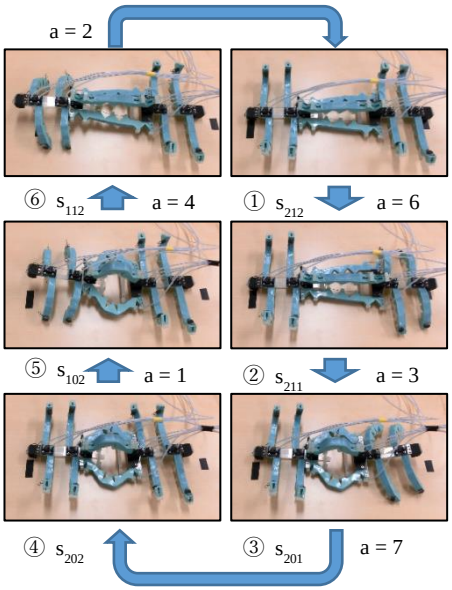


Fig.19 The obtained walking pattern

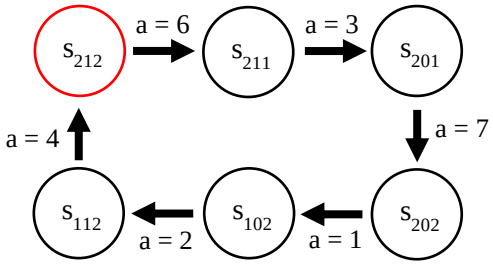


Fig.20 The state transition diagram (walking)

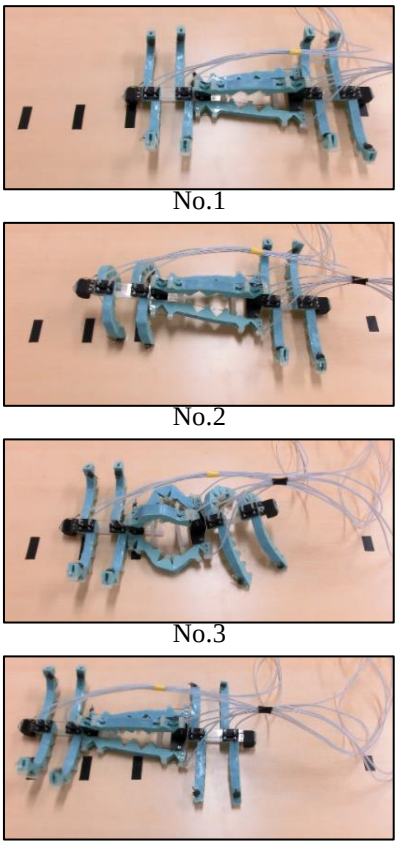


Fig.21 realized walking behavior

７．結論

本研究では，柔軟な身体を用いて強化学習の状態・行動空間を抽象化し，提案した枠組みを柔軟多脚型ロボットに適用した．ロボットが，パイプを登るという効果的な方策を容易に得られることを確認し，提案した枠組みが状態・行動空間の大きさ及び学習時間を減少させるのに効果的であることを実証した．

さらに，得られた方策は，追加学習なしに他の柱状物に適応可能であることを実験により確認し，汎化能力があることを示すことに成功した．

また，提案した枠組みの多様性を実証するため，異なる環境の例として，水平面で学習を行った．適切な歩行行動が得られたことから，提案した枠組みの汎用性が確認できた．

提案した枠組み（ソフトロボットと強化学習の組み合わせ）は効果的であり，有望であると結論付けることが出来る．今後は，この枠組みを実用的なものに応用したいと考える．

謝辞：最後に，本研究に際して多大なるご指導，ご協力をいただいた法政大学理工学部伊藤一之教授，伊藤研究室の方々，及び本研究の一部にご協力をいただいたイギリスのブリストルロボット研究所に心から感謝いたしております．また，今後の皆様のご健闘を願うとともに，法政大学における各研究において，本論文がほんのわずかながらでも参考になればと願うものであります．

参考文献

1) M. Hacibevoglu, A. Arslan,”Reinforcement learning accelerated with artificial network for maze and search problems”, *Proc. of 3rd International Conference on Human System Interaction*, pp. 124-127, 2010

2) M. Obayashi, K. Narita, Y. Okamoto, T. Kuremoto, K. Kobayashi, L. Feng, “A Reinforcement Learning System Embedded Agent with Neural Network-Based Adaptive Hierarchical Memory Structure”, *In Advances in Reinforcement Learning*, Chapter 11 , pp.189-208, IN-TECH , 2011

3) J. Asmuth, M. L. Littman, “Learning is planning: near Bayes-optimal reinforcement learning via Monte-Carlo tree search”, *arXiv preprint arXiv*: 1202.3699, 2012

4) N. A. Vien, W. Ertel, “Monte Carlo Tree Search for Bayesian Reinforcement Learning”, *Proc. of 11th International Conference on Machine Learning and Applications*, pp. 138-143, 2012

5) D. Zhao, H. Wang, K. Shao, Y. Zhu, “Deep reinforcement learning with experience replay based on SARSA”, *Proc. of 2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-6, 2016

6) K. Mo, H. Li, Z. Lin, J. Lee, “The AdobeIndoorNav Dataset: Towards Deep Reinforcement Learning based Real-world Indoor Robot Visual Navigation”, *arXiv preprint arXiv:* 1802.08824, 2018

7) M. Pfeiffer, S. Shukla, M. Turchetta, C. Cadena, A. Krause, R. Siegwart, J. Nieto, “Reinforced Imitation: Sample Efficient Deep Reinforcement Learning for Map-less Navigation by Leveraging Prior Demonstrations”, *arXiv preprint arXiv*: 1805.07095, 2018

8) N. A. Vien, W. Ertel, “Monte Carlo Tree Search for Bayesian Reinforcement Learning”, *Proc. of 11th International Conference on Machine Learning and Applications*, pp. 138–143, 2012

9) S. Seo, K. Ko, H. Yang, and K. Sim, “Behavior learning and evolution of swarm robot system using SVM”, *Proc. of 2007 International Conference on Control, Automation and Systems*, pp. 1238–1242, 2007

10) L. Chapel, G. Deffuant, “SVM Viability Controller Active Learning: Application to Bike Control”, *Proc. of 2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*, pp. 193–200, 2007

11) K. M. Digumarti, A. T. Conn, and J. Rossiter, “Euglenoid-inspired giant shape change for highly deformable soft robots”, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 494□501. 2017

12) K. Ito, Y. Fukumori, and A. Takayama, Autonomous control of real snake-like robot using reinforcement learning-abstraction of state-action space using properties of real world-, Proc. of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp.389–394, 2007.

13) G. Sumbre, Y. Gutfreund, G. Fiorito, T. Flash and B. Hochner, “Control of octopus arm extension by a peripheral motor program”, Science, Vol.293, No. 5536, pp. 1845–1848. 2001

14) Y. Gutfreund, T. Flash, G. Fiorito, and B. Hochner, “Patterns of arm muscle activation involved in octopus reaching movements,” Journal of Neuroscience, Vol. 18, No. 15, pp. 5976–5987. 1998.

15) C. J. C. H. Watkins and P. Dayan, “Q-learning”, Machine Learning, Vol. 8, pp. 279–292, 1992