

博士論文

報酬に基づく協調問題解決のための注目点の獲得に関する研究

京都工芸繊維大学 大学院工芸科学研究科
設計工学専攻

坂戸 達陽

平成 28 年 3 月 25 日

報酬に基づく協調問題解決のための注目点の獲得に関する研究

坂戸 達陽

概要

将来、各家庭にロボットが導入される日も遠くないと言われている。日常生活の場において人の手助けを目的とするロボットには、日常生活における様々な場面で起こる様々な状況に対応する能力が求められるが、起こり得るすべての状況を想定してロボットの動作を予め組み込んでおくことは困難である。よって、日常生活の場で活動するロボットには、インタラクションを通して、置かれている環境や他者に適応する能力が必要である。しかし、ロボットが特定の課題を達成する際、実世界の環境から得られる情報には、課題の達成とは関係のない情報も多い。また、複雑な処理を行うロボットは、内部で扱う記号間の関係性や、記号自体の意味を、環境に適した形で獲得する必要がある。さらに、様々な状況に対応するロボットは、特定の状況に対応するロボットよりも多くの機能を持っていると考えられるが、特定の課題を達成する際には必要とされない機能も多いはずである。よって、ロボットが実世界で学習する際には、環境から得られる情報から必要な情報を選択すること、そして、ロボットのもつ機能の中から、課題の達成に必要な機能を選択することが要求される。

本論文では、ロボットなどの学習エージェントが獲得する報酬に着目し、学習エージェントが課題の達成に必要な情報や機能などの要素を選択するための注目点を獲得するモデルを、(i) 外部情報、(ii) 内部表象、(iii) 内部処理の 3 つの観点から提案する。課題はそれぞれ、模倣学習、見立て遊び、協調行動という課題を用いる。課題には他者とのインタラクションを含め、協調問題解決のための注目点の獲得モデルを提案する。外部情報の注目点の獲得では、模倣学習における注目対象を獲得するモデルを提案する。ロボットなどの学習エージェントの学習手法の 1 つとして、試行錯誤による学習が挙げられるが、探索空間の広い環境においては、試行錯誤による学習のみでは、学習に多くの時間がかかり実用的ではない。そこで、

すでに達成すべきタスクに習熟している他者を模倣することで効率的に学習することが考えられるが、このような模倣学習における問題の1つとして、他者の行動のどの要素に注目するかということが挙げられる。本論文では、模倣学習において観測した他者の行動の何に注目するかという、模倣学習における注目対象を獲得するモデルを提案する。シミュレーション環境でのタスクとしては食卓上でのインタラクションを扱い、実世界のロボットでのタスクとしては塗り動作を扱う。実験の結果、提案モデルでの学習で、タスクに対して適切な注目対象を獲得することができた。

内部表象の注目点の獲得では、見立て遊びにおける学習エージェントの内的表象を獲得するモデルを提案する。知的エージェントのモデルの1つとして、認識、計画、実行などの機能ごとのモジュールの組み合わせによって複雑な問題を解決しようとするモデルがある。このモデルでは、モジュールの組み合わせを変えることで、異なる環境や状況に学習結果を再利用することができる。本論文では、インタラクションとして学習エージェントと他者との見立てを含む真似遊びを扱い、インタラクションを通して物体間の関係性を学習し、他者に伝わる見立てを行うための物体の表象を獲得するモデルおよび、各機能モジュール間の関係性を学習し、意図の表象を獲得するモデルを提案する。タスクとしては、見立ての獲得、他者に伝わる見立ての獲得、見立てを含む真似遊びにおける意図の表象の獲得を扱う。実験の結果、学習エージェントは、それぞれのタスクに対して提案したモデルで適切な学習を行うことができた。

内部処理の注目点の獲得では、他者意図の推定に基づく協調行動を獲得するモデルを提案する。ロボットが人の手助けを目的とする場合、人の意図を推測し、推測した意図に基づいた協調行動を行う能力が必要である。本論文では、モジュール型のモデルを用いた、他者意図の推定に基づく協調行動を獲得するモデルを提案する。タスクとしては積み木遊びを扱い、モジュールの組み合わせによって協調行動を獲得するモデルを提案する。実験の結果、学習エージェントは与えられたタスクにおける協調行動に必要なモジュールの組み合わせ方を学習し、他者と協力してタスクを達成することができた。

Research on the Acquisition of Attention Points for Cooperative Problem Solving Based on Rewards

SAKATO Tatsuya

Abstract

In the near future, robots are expected to be introduced to each home. In the field of daily life, the robots for helping people need to respond to various situations in daily life. However, it is virtually impossible to predict all possible situations and to preprogram the robot with all suitable behavioral patterns for each of the possible situations. Thus, the robot needs to learn suitable behavior through interaction. However, when the robot is to achieve a particular task, information obtained from the real world environment, there is much information unrelated to accomplish the task. Also, a robot which is able to adapt to various circumstances has many functions which are not needed for achieving a particular task. Thus, when the robot learns in the real world environment, it has to select necessary information from information obtained from the environment, and, necessary functions among the functions of the robot.

This thesis focuses on acquisition of attention points for cooperative problem solving based on rewards received by a learning agent. Three key studies each of which focused on (i) external information, (ii) internal symbols, or (iii) internal processing are discussed.

First, acquisition of attention points of observed actions in an imitation task is discussed as acquisition of attention points of external information. The experimental results indicate that the proposed model can acquire suitable attention points of external information through the task. Second, acquisition of symbols of intentions in pretend play is discussed as acquisition of attention

points of internal symbols. The experimental results indicate that the proposed model can acquire suitable attention points of internal symbols through the task. Finally, achieving cooperative behavior based on intention estimation by learning combinations of modules is discussed as acquisition of attention points of internal processing. The experimental results indicate that the proposed model can acquire suitable attention points of internal processing through the task.

Through those three key studies, acquisition of attention points for cooperative problem solving based on rewards is discussed.

謝辞

本論文は筆者が京都工芸繊維大学 大学院工芸科学研究科 博士後期課程 設計工学専攻に在籍中の研究成果をまとめたものである。同大学教授 岡 夏樹先生には、指導教官として、本研究を実施する機会を与えていただき、また、その遂行および論文執筆にあたって多大なるご指導をいただいた。ここに感謝の意を表する。

同大学准教授 荒木 雅弘先生には研究および論文執筆に関して多大なるご指導をいただいた。ここに感謝の意を表する。

また、インタラクティブ知能研究室の各位には、研究遂行にあたり日頃より有益なご討論ご助言をいただいた。ここに感謝の意を表する。

最後に、これまで私をあたたく応援し励ましてくれた、家族と友人に心より感謝の意を表する。

参考文献

- [1] A. Oliva, A. Torralba, M.S. Castelhano and J.M. Henderson. “Top-down control of visual attention in object detection.” in *IEEE Proc. the International Conference on Image Processing*, 2003, pp. 253-256.
- [2] V. Navalpakkam and L. Itti. “*Modeling the influence of task on attention.*” Vision Research, vol. 45, pp. 205-231, 2005.
- [3] A. Alissandrakis, C.L. Nehaniv and K. Dautenhahn. “Imitation with alice: Learning to imitate corresponding actions across dissimilar embodiments.” in *IEEE Transactions on Systems and Humans*, vol. 32, 2002, pp. 482-496.
- [4] A. Billard, Y. Eparsm S. Calinon, S. Schaal and G. Cheng. “*Discovering optimal imitation strategies.*” Robotics and Autonomous Systems, vol. 47, pp. 69-77, 2004.
- [5] A. Zook, B. Magerko and M. Riedl. “Formally modeling pretend object play.” in Proc. the 8th ACM Conference on Creativity and Cognition, 2011, pp. 147-156.
- [6] B. Magerko, J. Perner, M. Jacob, M. Comerford and J. Smith. “An overview of computational co-creative pretend play with a human.” in *Proc. the Playful Virtual Characters Workshop at the Fourteenth Annual Conference on Intelligent Virtual Agents 2014*, 2014.
- [7] T. Araki, T. Nakamura, T. Nagai, K. Funakoshi, M. Nakano and N. Iwahashi. “Autonomous acquisition of multimodal information for online object concept formation by robots.” in *Proc. the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 1540-1547.
- [8] 小川 昭利, 大森 隆司. “機能部品組み合わせモデルによるナビゲーション行動学習処理の獲得方式の提案.” 電子情報通信学会論文誌, vol. J87-D-II, 2004, pp. 287-998.
- [9] N. Oka. “Apparent “free will” caused by representation of module control,” in *No matter, Never mind: Proc. Toward a Science of Consciousness: Fundamental Approaches*, 1999, pp. 243-249.

- [10]J.C. Sylvester, J.A. Reggia, S.A. Weems and M.F. Bunting. “Controlling working memory with learned instructions.” *Neural Networks*, vol. 41, pp. 23-38, May 2013.
- [11]R.S. Sutton and A.G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 1998.
- [12]E. Wiewiora, G. Cottrell and C. Elkan. “Principled methods for advising reinforcement learning agents.” in *Proc. the Twentieth International Conference on Machine Learning*, 2003, pp. 792-799.
- [13]A.Y. Ng, D. Harada and S. Russell. “Policy invariance under reward transformations; Theory and application to reward shaping.” in *Proc. the Sixteenth International Conference on Machine Learning*, 1999, pp. 278-287.
- [14]B. Price and C. Boutilier. “Implicit imitation in multiagent reinforcement learning.” in *Proc. the Sixteenth International Conference on Machine Learning*, 1999, pp. 325-334.
- [15]田淵 一真, 谷口 忠大, 榎木 哲夫. “模倣学習と強化学習の調和による効率的行動獲得.” 2006年度人工知能学会全国大会 (第20回) 論文集, 2006.
- [16]C.L. Nehaniv and K. Dautenhahn. “*Of humming birds and helicopters: An algebraic framework for interdisciplinary studies of imitation and its applications.*” INTERDISCIPLINARY APPROACHES TO ROBOT LEARNING, 1999, pp. 136-161.
- [17]H. vanHasselt. “Reinforcement learning in continuous state and action spaces.” in *Reinforcement Learning: State of the Art*, Springer, 2012, pp. 207-252.
- [18]久崎 孝浩. “生後2年目における認知発達 —表象機能という視点からの考察—.” 九州大学心理学研究, vol. 4, 2003, pp. 37-55.
- [19]志波 泰子. “2歳児は誤信念を理解するだろうか: PernerとLeslieの論争を再考する.” 京都大学大学院教育学研究科紀要, vol. 55, 2009, pp. 75-87.
- [20]井上 洋平, “幼児期におけるふり行動の発達の研究 —ふり行動の二重性に関する

- る一考察一.” 立命館産業社会論集, vol. 43, 2007, pp. 77-93.
- [21] R. Jacobs, M. Jordan, S. Nowlan and G. Hinton. “Adaptive mixture of local experts”. *Neural Computation*, vol. 3, pp. 79-87, 1991.
- [22] 坂本 裕太, 坂戸 達陽, 尾関 基行, 岡夏樹. “モジュール組換え型モデルにおけるモジュールの学習とモジュール組換え系列の学習.” 2012 年度人工知能学会全国大会 (第 26 回) 論文集, 2012, 3B2-R-2-6.
- [23] S. Ikenooue, M. Asada and K. Hosoda. “Cooperative behavior acquisition by asynchronous policy renewal that enables simultaneous learning in multiagent environment,” in *Proc. the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, 2002, pp. 2728-2734.
- [24] T. Makino and K. Aihara. “Multi-agent reinforcement learning algorithm to handle beliefs of other agents’ policies and embedded beliefs,” in *Proc. International Joint Conference on Autonomous Agents and Multiagent Systems*, 2006, pp. 789-791.
- [25] Y. Nagayuki, S. Ishii and K. Doya. “Multi-agent reinforcement learning: An approach based on the other agent’s internal model,” in *Proc. Fourth International Conference on MultiAgent Systems*, 2000, pp. 215-221.
- [26] K. Kobayashi, K. Nakano, T. Kuremoto and M. Obayashi. “Cooperative behavior acquisition of multiple autonomous mobile robots by an objective-based reinforcement learning system,” in *Proc. International Conference on Control, Automation and Systems*, 2007, pp. 777-780.