

令和 4 年 6 月 14 日現在

機関番号：32657
 研究種目：基盤研究(C)（一般）
 研究期間：2019～2021
 課題番号：19K12173
 研究課題名（和文）プロトタイプ理論に基づく強化学習ロボットの知識選択における認知的経済性の向上
 研究課題名（英文）Cognitive Economy in Reusing Policy Selection for Reinforcement Learning Robots Based on Prototype Theory
 研究代表者
 鈴木 剛（Suzuki, Tsuyoshi）
 東京電機大学・工学部・教授
 研究者番号：00349789
 交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：転移学習を用いた強化学習ロボットの認知的経済性を実現するため、学習知識のカテゴリ化とプロトタイプの抽出、知識選択の高速化について検討した。最短経路探索を対象に、活性化拡散モデルに基づく知識のネットワーク化、プロトタイプ理論に基づくK-means++を用いた知識のカテゴリ化、カテゴリ内知識の平均化によるプロトタイプ抽出を行い、計算機実験により学習時間の短縮を確認した。計算機クラスターを用いた並列計算による知識選択時の計算高速化を図り、自律移動ロボットに実装し効果を検証した。物体形状のカテゴリ化とプロトタイプ抽出のために、学習による物体のプリミティブ形状認識を行い、形状に適した物体操作を実現した。

研究成果の学術的意義や社会的意義
 本研究の最終的な目標は、「直観」や「直感」といったヒトの無意識的な判断機構をモデル化し、学習ロボットに実装することである。これにより、知能ロボットの新たなタスクへの導入時に、環境の拘束条件の緩和、タスク適応的な行動の迅速な獲得、咄嗟の環境条件変化への対応などが期待できる。本研究課題では、その基礎検討として、認知心理学の知見である活性化拡散モデルおよびプロトタイプ理論の導入による強化学習ロボットの効率的かつ高速な知識のカテゴリ化と選択（認知的経済性）の方法について提案し、実験により検証した。本研究課題の成果は、学習ロボットや学習エージェントの実用化・普及への貢献が期待できる。

研究成果の概要（英文）：In order to realize cognitive economy of reinforcement learning robot using transfer learning, we studied categorization of reusing learning policies, extraction of prototypes in category, and speed-up of reusing policy selection. For shortest path search problem, we performed networking of reusing policies based on spreading activation model, categorization of policies using K-means++ based on prototype theory, and extraction of prototypes by averaging policies within a category, and confirmed the reduction of learning time through computer experiments. We also performed parallel computation using computer clusters for speed-up of computation during policy selection, and verified the effectiveness by implementing the method on an autonomous mobile robot. For object shape categorization and prototype extraction, primitive shape recognition was performed by learning, and shape-appropriate object manipulation was executed.

研究分野：ネットワークロボット

キーワード：転移学習 システム 認知心理学モデル マルチエージェント強化学習 機械学習 強化学習 マルチエージェントロボット

1. 研究開始当初の背景

深層学習などの機械学習手法や機械学習アルゴリズムの発展に伴い、それらを用いた人工知能や知能ロボットの様々な分野、様々なタスクへの応用が期待されている。学習アルゴリズムの実世界応用を考慮した場合、ある環境での学習で得た特定の知識のみで多種多様なタスクに適応することは難しく、様々な環境で学習した知識群から適宜必要な知識(群)を選択して振る舞う方が技術的にも現実的と考える。しかし、このような知識の選択手法に関してはほとんど議論されていない。

申請者らは、転移学習を用いた強化学習を基盤に、認知心理学から得られた知見である、ヒトの無意識的な概念の想起方法である活性化拡散モデルを用いた強化学習ロボットにおける知識選択手法を提案し、評価してきた(図1)。活性化拡散モデルとは、人間が獲得した概念同士が脳内でネットワーク構造として保存されていることを前提に、ある概念が想起されることで関連する概念も活性化され、概念の利用が促進されるモデルである^[1]。これにより、与えられたタスクと関連性の高い知識の想起に基づいて知識の選択と利用が可能であることを確認したが、知識量が膨大になると知識選択のための計算量も指数関数的に増大することが明らかとなった。ヒトの認知行動では知識選択においてこのような計算量の増大は発生しない(認知的経済性)ことから、本研究課題では、認知的経済性を強化学習ロボットにより実現する手法について研究する。なお、ここでの知識とは、ある環境でタスクを実行する際の刺激(センサ入力など)に対する反応(行動の方策など)の集合とする。

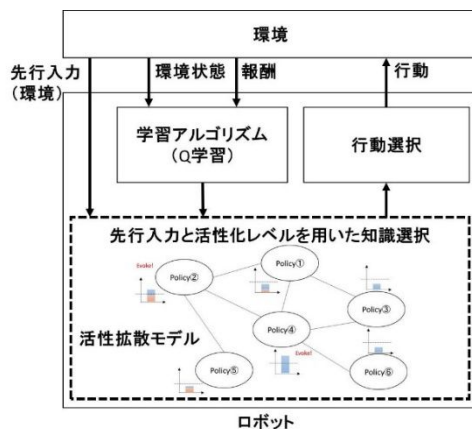


図1 活性化拡散モデルを用いた知識選択

2. 研究の目的

申請者らの研究の最終的な目標は、「直観(intuition)」や「直感(inspiration)」といったヒトが無意識的に行う判断のメカニズムをモデル化し、学習ロボットに実装することである。「直観」は論理的な、「直感」は感覚的なひらめきと考えられ、それぞれ意味は異なるが、いずれも既獲得の知識や経験に基づき、ヒトが無意識的(思考や分析の時間を経ず)に問題に対する適応的な対応を判断するものと考えられる。これを実装することで、例えば、知能ロボットの新たなタスクへの導入において、環境の拘束条件の緩和、タスクに適応的な行動の迅速な獲得、環境条件の変化への咄嗟の対応などが期待できる。本研究課題では、そのアプローチの一つとして、主として「直観」的な判断機構の実装を目指し、これまで検討してきた活性化拡散モデルに加え、プロトタイプ理論の導入による強化学習ロボットの効率的かつ高速な知識のカテゴリ化と選択(認知的経済性)の実現を目的とする。すなわち、強化学習ロボットが、既獲得の複数の知識(知識群)をプロトタイプ理論に基づき分類し、与えられたタスクの達成に必要な知識群を活性化拡散モデルに基づき素早く選択する手法を構築する。

3. 研究の方法

認知心理学におけるプロトタイプ理論は、ヒトが見たモノや言語的な特徴量(概念)を、距離を用いて分離(カテゴリ化)して、カテゴリ内の概念を代表する同じ傾向のまとめ(プロトタイプ)として表現し、ヒトが何かを認知する際にはプロトタイプとの類似性を用いて高速に認知(認知的経済性を実現)しているとするものである^{[2]他}。このため、ヒトは全ての知識や概念を検索せず、効率良く物事の認知が可能と考えられる。

本研究課題では、図2のように従来の知識ネットワーク内の知識群を特徴量に基づきカテゴリとして分類し、仮想的なノードとしてカテゴリ内の知識の傾向を表すプロトタイプを知識ネットワークに接続する。知識の選択では、観測された環境特徴量からプロトタイプを選択し、プロトタイプに接続された知識を一斉に活性化することで、知識選択時の計算量削減を目指す。

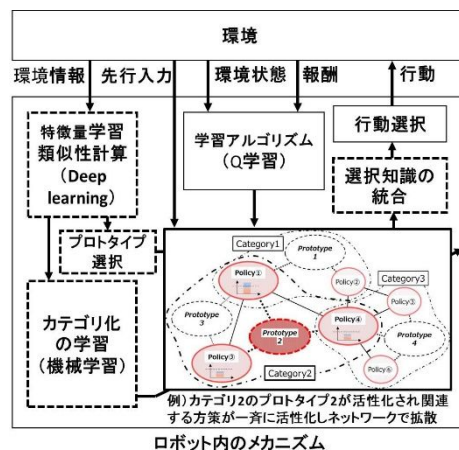


図2 提案手法の概念

4. 研究成果

本研究課題では、活性化拡散モデルおよびプロトタイプ理論に基づき、転移学習を用いた強化学習ロボットの認知的経済性を実現するために、学習知識のカテゴリ化とプロトタイプ抽出の

方法、活性化拡散モデルにおける知識選択の高速化を図る手法などについて検討した。

4.1 プロトタイプ理論に基づきカテゴリ化した知識を用いた転移学習

複数の知識に基づく転移学習を実現するために、活性化拡散モデルとプロトタイプ理論を導入した転移学習手法を提案した。ここではQ学習を使用し、ある環境での学習で得られたQテーブルを一つの知識としている。

活性化拡散モデルは知識間距離ネットワークで表現する。このネットワークを本研究では、Spreading Activation Policy Network (SAP-Net)と呼び、転移学習中の手続きに利用する^[3]。SAP-Netでは知識の類似度によって知識間距離を設定する。本提案において知識間距離は、各Qテーブルの行動価値を比較したユークリッド距離を正規化することで求める。次に、プロトタイプ理論を導入するために知識をK-means++^[4]を利用してカテゴリ化する。分類点と各知識との距離は、SAP-Netの知識間距離に基づく。これを任意の回数行い平均シルエット係数^[5]が最も高い分類結果をカテゴリとして採用する。これにより、知識間の接続とカテゴリ化を行い、さらに作成された各カテゴリに属する知識の行動価値を平均化し、カテゴリの知識の傾向を表す「プロトタイプ知識」を生成する。

学習エージェント(強化学習ロボットなど)は、新たなタスク遂行時に、プロトタイプ知識の試行と評価を繰り返すことで、現在のタスクに適したカテゴリを特定する。そして、特定したカテゴリの中から、活性化拡散モデルの活性値を基に知識を選択し転移する。

最短経路問題を学習タスクとして、計算機実験により提案手法による学習時間を評価した。学習エージェントは、図3に示すフィールド内にスタート地点とゴール地点を設け、これらをつなぐ経路が最短となる順路を学習する。ここでは、フィールドの大きさを11×11マスの離散環境とし、1マスの移動などの1回の行動を1[step]とした。図のSがスタート、Gがゴール、白マスが通路、黒マスが障害物である。行動機構は前、後、右、左それぞれ1マスの移動と、その場のマスでの停止の5つである。実験は、最短経路学習を、強化学習、活性化拡散モデル単体、提案手法(活性化拡散モデル+プロトタイプ理論)でそれぞれ

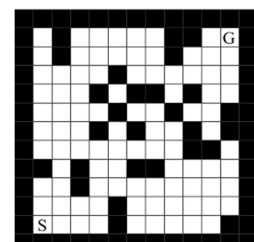


図3 経路探索環境

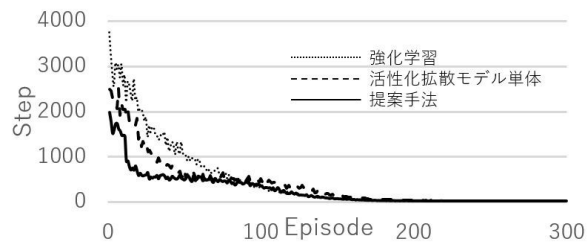


図4 episode毎のstep数の推移

を行い、総step数(学習エージェントの総移動回数)を比較した。なお、活性化拡散モデル単体と提案手法においては、スタート地点とゴール地点、および、障害物配置が全て異なる環境で予め学習した100個の知識を用意し、新たな環境における学習時にその知識群から適した知識を転移する。提案手法のカテゴリ数は2~20で行い、最も総ステップ数の少ないカテゴリ数で比較した。実験は200回試行し、その平均値を評価した。1試行は300episodesである。

実験の結果、総step数の平均は、強化学習143073.77[step]、活性化拡散モデル単体117185.16[step]となった。提案手法での総step数平均はカテゴリ数により異なり、カテゴリ数18で最小の77621.21[step]となった。これより、提案手法は、強化学習に対して総ステップ数平均が約46%、活性化拡散モデル単体に対して約34%減少した。学習効率の向上が見られることから、提案手法により、新たなタスクに適した知識が選択されていることが分かる。また、提案手法と他学習手法のepisode毎のstep数の推移を図4に示す。1~100episodeにおいて提案手法は他手法より最大step数が39564.16[step]減少しており、タスクに適した知識を学習初期から選択できている。これらのことから、プロトタイプ理論を用いた提案手法により、活性化拡散モデルによる転移学習において、効率的かつ高速な知識選択(認知的経済性)に対する効果が見られた。

4.2 並列計算を用いた活性化拡散モデルに基づく知識選択の高速化

SAP-Netでは、実装を行う場合に選択可能な知識数が増加すると知識選択を行う計算量が指数関数的に増大し、シーケンシャルな知識選択実装では計算時間の課題があった。そこで、各知識の活性化計算を並列化することで計算量の低減、すなわち認知的経済性に寄与する新たなSAP-Netの実装手法を実現した。本来ヒトの脳内で動作しているとされる活性化拡散モデルの動作では問題とならない計算量リソースの制約を緩和するための手法を提案し、具体的にはSAP-Net中の各知識における活性値計算を、計算機クラスタを構成する各計算ノードに割り当て、同時並列的に活性値計算、拡散計算、減衰計算を行う。また、本研究では計算量を低減した計算機クラスタにより実装されたSAP-Netを、実

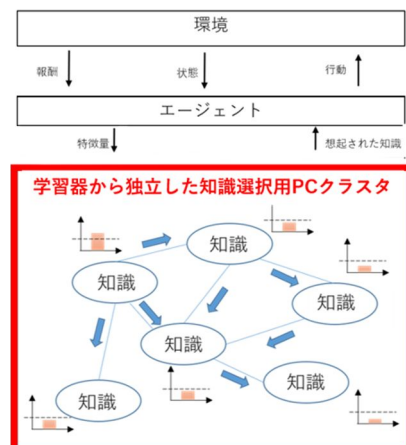


図5 活性化計算の高速化

ロボットプラットフォームに実装し実環境での知識選択の有用性を評価した。

本提案手法では、学習エージェントや学習ロボット内に実装される SAP-Net のシステム構成を、図 5 に示すように SAP-Net の計算処理部分を切り離し計算機クラスタによる並列計算により実現した。従来通り強化学習機能や転移学習機能は学習エージェントや学習ロボットのエッジ側で動作し、通信インフラを利用して知識選択処理や選択された知識の転送機能をサーバシステム(計算機クラスタ)と連携する。計算機クラスタは Raspberry Pi 3 model B+を 40 台搭載し、各 Raspberry Pi には 4 コアの CPU が搭載されているため、合計 160 コアを提供可能である。計算機クラスタの制御には並列コンピューティングを行うためのライブラリ MPI (message passing interface)を使用し、Python により SAP-Net の並列化を実装した。学習エージェントや学習ロボットと計算機クラスタの通信にはインターネット回線を利用し VPN (virtual private network)を利用することで End-to-end 間の通信を確立している。

提案手法の評価を実ロボットと実環境で評価するために、図 6 に示す自律運転シニアカーを開発し実験を行った。自律運転シニアカーはスズキ株式会社製のシニアカーである ET4D にセンサやアクチュエータ類を搭載し自律運転を可能としているレトロフィット型である。センサとしてエージェント周囲の環境情報の取得用として Velodyne 社製 3 次元 LiDAR VLP-16 が搭載されている。また、転移学習時に必要となる自己位置座標と姿勢角の取得用に Intel Realsense T265、本実験における SAP-Net を用いた転移学習において環境特徴量となるデータの取得を行うための北陽電機社製測域センサ LRF UST-10LX が合わせて搭載されている。また、自律運転シニアカー以外にも、マルチエージェントシステムにおける提案手法評価用に小型フォークリフト型ロボットを開発した。

実験により、自律運転シニアカーが実環境で走行するための知識を選択し、障害物を回避しながら走行できるかを確認した。あらかじめ複数環境の形状情報を LiDAR により点群データで取得し、物理演算シミュレータである Webots による仮想シニアカーモデルの移動経路強化学習を行う。学習により得られた複数の知識を計算機クラスタで SAP-Net を用いて管理し、走行環境に適した知識を実環境のシニアカーに転移し走行、障害物回避を実現する。

実験結果として、自律運転シニアカーがオンサイトで環境情報を取得し、計算機クラスタを用いて知識選択、実行する知識のダウンロードを行い実環境で走行が可能であることを確認した。図 7 に移動軌跡の比較を示す。移動経路に累積誤差が見られるものの事前の強化学習シミュレーションにて獲得した移動経路と、実際に自律運転シニアカーが走行した経路が同じ傾向にあり、なおかつ自律運転シニアカーが障害物を認識し計算機クラスタから正しく知識を選択できたことが確認できた。さらに、自律運転シニアカーの走行中に障害物が出現し、障害物のない環境における直進走行の知識から障害物を回避するための走行経路の知識に動的に知識選択する実験を行い、自律運転シニアカーの走行中における知識選択を実現し、本提案手法における計算量低減の効果が表れていることを確認した。

4.3 プリミティブ物体のパレット輸送における深層学習を用いた位置姿勢認識と積載安定化

実環境における物体のハンドリングタスク等を考えた場合、物体形状が持つ本質的な形(プリミティブ形状)を認識することで、多様な物体をプリミティブ形状でカテゴリ化し、形状に共通するプロトタイプを抽出できる。また、形状に対応した物体操作を学習することで、物体形状に適した操作知識の選択が可能となると考えられる。そこで、学習により物体のプリミティブ形状を認識し、形状に適した物体のハンドリングを行う方法について検討した。

2020年度には、直方体や球などで表されるプリミティブ形状対象を認識し、作業主体である学習エージェントに対する位置姿勢を認識するための判定器を、機械学習における深層学習の技術で構築した。具体的には RGB-D センサを用いて、画像データからプリミティブ形状物体を検出し、その領域に対



図6 自律運転シニアカー

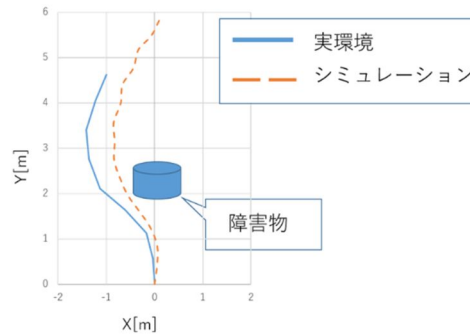


図7 実験結果の移動軌跡

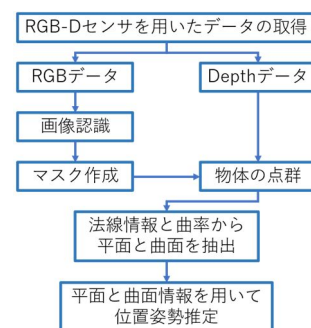


図8：位置姿勢推定フロー

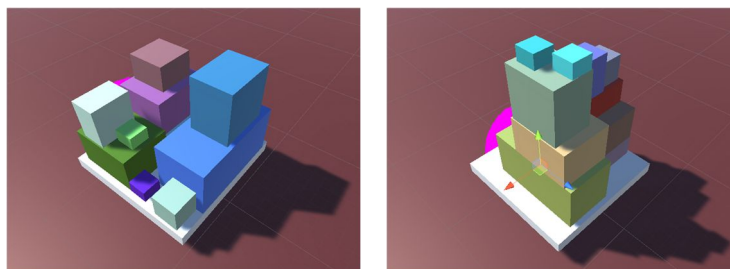
応する3次元物体の位置姿勢を推定するシステムを構築した(図8)。プリミティブ形状の検出には、画素単位で対象を検出可能なMask R-CNN^[6]を利用し、検出したプリミティブ形状(図9)の位置姿勢推定は独自の幾何学モデルを用いて行った。



(a) 認識対象 (b) 検出結果

図9：プリミティブ形状の位置姿勢推定

2021年度には、上記の技術で検出されたプリミティブ形状の搬送問題に取り組んだ。認識した複数の対象を安定して輸送するためには、各物体を適切な順番で適切な位置に積載する必要がある。本課題では、この問題を小型フォークリフト型ロボットによる荷物輸送問題として定式化した。大きさの異なる複数の直方体に対して、輸送中のカーブなどの回転時に生じる遠心力の影響を考慮して、安定して輸送するための積載方法を獲得するアルゴリズムを提案した。具体的には、積載物全体の安定性を評価するための定量的な指標を提案し、積載過程における不安定さを許容することで全体的な安定性を向上させることが可能な探索型の積載パターン生成アルゴリズムを構築した。提案した安定性の評価指標に基づく安定な積載例と、不安定な積載例を図10(a)、図10(b)にそれぞれ示す。安定性の高い図10(a)は、安定性の低い図10(b)よりも、重心が低く、水平方向に荷物が展開していることから回転時の遠心力による外向き方向の外力により頑健であることが確認できる。実験では、物理シミュレーションによって、提案した安定性の評価指標の有効性と、獲得された積載パターンの安定性を検証した。



(a) 安定した積載例 (b)不安定な積載例

図10：獲得した積載パターン

4.4 まとめと今後の展望

本研究課題では、強化学習ロボットの認知的経済性を実現するために、ヒトの無意識的な判断機構をモデル化し強化学習ロボットに実装することを目標に、その基礎検討として認知心理学の知見である活性化拡散モデルおよびプロトタイプ理論の導入による強化学習ロボットの効率的かつ高速な知識のカテゴリ化と選択の方法について提案し、実験により検証を行った。今後、本成果を実環境における複数のタスクを対象として実ロボットに統合し、効果の検証を行う予定である。本研究課題の成果は、強化学習以外の学習アルゴリズムでも適応可能であり、学習ロボットや学習エージェントの実用化・普及への貢献が期待できると考える。

<引用文献>

- [1] A. M. Collins: "A Spreading-Activation Theory of Semantic Processing", *Psychological review*, Vol. 82, No.6, pp.407-427(1975)
- [2] 安藤清志, 石口哲, 高橋晃, 浜村良久, 藤井輝男, 八木保樹, 山田一之, 渡邊正孝, 重野純, "心理学 改訂版", 株式会社新曜社, 978-4-7885-1290-0, (2012)
- [3] H. Kono, R. Katayama, Y. Takakuwa, W. Wen and T. Suzuki, "Activation and Spreading Sequence for Spreading Activation Policy Selection Method in Transfer Reinforcement Learning" *International Journal of Advanced Computer Science and Applications (IJACSA)*, 10(12), pp.7-16, (2019)
- [4] D. Arthur and S. Vassilvitskii: "k-means++: The Advantages of Careful Seeding", *SODA*, pp.1027-1035 (2007)
- [5] P. J. Rousseeuw: "Silhouette: a graphical aid to the interpretation and validation of cluster analysis", *J. of computational and applied mathematics*, Vol.20, pp.53-65(1987)
- [6] K. He, G. Gkioxari, P. Dollár, R. Girshick, "Mask R-CNN," In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961-2969, (2017)

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Yusaku Takakuwa, Wen Wen, Hitoshi Kono, Tsuyoshi Suzuki, Hiromitsu Fujii	4. 巻 12
2. 論文標題 Autonomous Reusing Policy Selection using Spreading Activation Model in Deep Reinforcement Learning	5. 発行年 2021年
3. 雑誌名 International Journal of Advanced Computer Science and Applications	6. 最初と最後の頁 pp.8-15
掲載論文のDOI（デジタルオブジェクト識別子） 10.14569/IJACSA.2021.0120402	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Hitoshi Kono, Ren Katayama, Yusaku Takakuwa, Wen Wen, Tsuyoshi Suzuki	4. 巻 10
2. 論文標題 Activation and Spreading Sequence for Spreading Activation Policy Selection Method in Transfer Reinforcement Learning	5. 発行年 2019年
3. 雑誌名 International Journal of Advanced Computer Science and Applications	6. 最初と最後の頁 pp.7 - 16
掲載論文のDOI（デジタルオブジェクト識別子） 10.14569/IJACSA.2019.0101202	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計5件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 坂本裕都, 河野仁, 温文, 藤井浩光, 鈴木剛
2. 発表標題 転移学習を用いた強化学習ロボットの方策選択における認知的経済性の検討
3. 学会等名 日本機械学会ロボティクス・メカトロニクス講演会2020
4. 発表年 2020年

1. 発表者名 野口達矢, 藤井浩光
2. 発表標題 深層学習を用いた物体抽出と距離情報を用いた面検出によるプリミティブ形状の位置姿勢推定
3. 学会等名 日本機械学会ロボティクス・メカトロニクス講演会2020
4. 発表年 2020年

1. 発表者名 坂本裕都、河野仁、温文、藤井浩光、鈴木剛
2. 発表標題 転移強化学習のためのPCクラスタを用いた再利用知識選択アルゴリズムの開発
3. 学会等名 2021年電気学会電子・情報・システム部門大会
4. 発表年 2022年

1. 発表者名 秋山 智暉、河野 仁、温 文、藤井 浩光、鈴木 剛
2. 発表標題 プロトタイプ理論に基づきカテゴライズした知識を用いた転移学習
3. 学会等名 第22回計測自動制御学会 システムインテグレーション部門講演会
4. 発表年 2021年

1. 発表者名 金浪 由宇、藤井 浩光
2. 発表標題 パレット輸送における荷崩れ防止のための旋回動作中の遠心力を考慮した積載安定化
3. 学会等名 日本機械学会ロボティクス・メカトロニクス講演会2022
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>〔大学紀要〕</p> <p>(1) 朝倉隆道, 河野仁, 人工知能研究における工学と社会学との融合の可能性と課題 研究開発と社会的期待との関係に対する一考察, 東京工芸大学工学部紀要 人文・社会編, vol.43, no.2, pp. 21-26, 2020.</p> <p>(2) 朝倉隆道, 河野仁, 日本におけるプログラミング教育の受容 テクノロジーと社会的期待との乖離に関する一考察, 東京工芸大学工学部紀要 人文・社会編, vol.44, no.2, pp. 31-36, 2021.</p>
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	藤井 浩光 (Fujii Hiromitsu) (30781215)	千葉工業大学・先進工学部・准教授 (32503)	
研究分担者	温 文 (Wen Wen) (50646601)	東京大学・大学院工学系研究科(工学部)・特任准教授 (12601)	
研究分担者	河野 仁 (Kono Hitoshi) (70758367)	東京工芸大学・工学部・准教授 (32708)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関