

知識選択型転移強化学習を用いたシニアカーの自律運転

河野 仁* 坂本 裕都 (東京工芸大学)

温 文 (東京大学) 藤井 浩光 (千葉工業大学)

池 勇勲 (北陸先端科学技術大学院大学) 鈴木 剛 (東京電機大学)

Autonomous Driving of Senior Car using Transfer Reinforcement Learning with Knowledge Selection

Hitoshi Kono*, Yuto Sakamoto, (Tokyo Polytechnic University)

Wen Wen, (The University of Tokyo)

Hiromitsu Fujii, (Chiba Institute of Technology)

Yonghoon Ji, (Japan Advanced Institute of Science and Technology)

Tsuyoshi Suzuki, (Tokyo Denki University)

In recent years, next-generation transportation is expected due to reasons such as the return of driver's licenses and super-aging society. Senior cars that do not require a driver's license have been developed and are becoming widespread, however traffic accidents have occurred due to deterioration of the motor and cognitive functions of the elderly. In this paper, it is introduced that a system in which LiDAR is installed in a senior car, reinforcement learning of the driving route is virtually executed from the measured data with a physical calculation simulator, and the learning result is transferred to an actual senior car to realize autonomous driving. Furthermore, the result is explained that when an object such as a pedestrian appears on the driving route, the knowledge of running while avoiding object is automatically selected and an autonomous driving can be adapted to the environment.

キーワード：転移強化学習, Sim-to-Real, 知識選択, シニアカー, 自動運転

(Keywords, transfer reinforcement learning, Sim-to-Real, senior car, autonomous driving)

1. はじめに

超高齢化社会の到来を背景とする高齢者の運転免許返納などにより、次世代の移動手段が注目されている。次世代の移動手段としては、自動車の自動運転などが研究開発されているが、パーソナルモビリティとしては寸法や価格など一般利用者の入手性を考慮するとシニアカーのような規模が望ましいと考えられる。シニアカーは運転免許が不要なため、免許返納後の高齢者の移動手段として期待されている。しかし、シニアカーは高機動性ゆえに側溝や不整地走行での転倒事故などが懸念される。そのため、シニアカーの運転が難しい搭乗者には自動運転などの支援システムが事故発生率の低減に貢献できると考えられる。浅田や石田、沈らなどの研究では自律走行シニアカーが開発され、自律走行のための経路計画やセンサ活用などの研究が行われている⁽¹⁾⁽²⁾⁽³⁾。しかし、平坦で整地された経路などでは有用であるが公園内の走行や自宅敷地内、地図に無い近道など事前に環境情報の取得が難しい環境は考慮されていない。一方、河

野らはクローラ型災害対応ロボットの自律走行を目的としてオンサイトで取得した環境情報から走行経路を仮想的な物理演算シミュレーションにて強化学習し、学習した経路を実ロボットで実行するシステムを開発している⁽⁴⁾。しかし、ロボットや移動プラットフォームが様々な環境に即座に適応するための議論は十分になされていない。そこで、本研究では自律運転シニアカーの実現を目的として、シニアカーが環境で取得した環境情報ととりわけ LiDAR (light detection and ranging) による 3 次元形状情報を基に物理演算シミュレーションを用いて走行経路を学習、学習した結果にもとづき自律的に実シニアカーが走行、さらには学習した結果を蓄積、選択して再利用することで様々な環境で即座に自律走行するシニアカーシステムを提案する。本論文では自律運転シニアカーを開発し、LiDAR から取得した 3 次元形状から走行経路を強化学習、学習した知識を選択するための計算機クラスタを構築したので報告する。また、評価実験から自律走行シニアカーの進行方向に歩行者などの障害物が存在する場合を想定し、直線走行の知識(学

習済走行経路) から障害物回避の走行を行う知識へ自動的に切り替える知識選択の動作を確認した。

2. 関連研究とアプローチ

〈2・1〉 知識選択アルゴリズム これまで機械学習とリわけ強化学習において学習済みの知識を蓄積・選択して環境に適応的な振る舞いを実現する手法が提案されている。例えば PRQ-Learning や Strategy P などが提案されている⁽⁵⁾⁽⁶⁾。ここで述べる知識は強化学習における方策や行動価値関数を想定している。しかしこれらの研究は単一システムが選択可能な知識数や複雑な環境での検証など、検討されていないことも多い。Kono らや Takakuwa らなどは、心理学におけるヒトの脳内の活性化拡散モデルを強化学習の知識選択に応用した SAP-net を提案している⁽⁷⁾⁽⁸⁾。システム構成としては知識数に依存しない知識選択が可能なシステムであるが、シーケンシャルに処理するようなプログラミング実装では検索時間が方策数に依存する課題がある。しかし、SAP-net はスケーラビリティの高い手法であるため、本研究では知識選択手法として SAP-net を採用する。

〈2・2〉 Sim-to-Real 機械学習などのシミュレーションでは、現実環境との乖離から学習結果が現実世界で使用することが難しい課題に対して、近年はシミュレーションの環境をなるべく現実世界に近づけて学習し、その学習結果を実ロボットで実行する Sim-to-Real というフレームワークが研究されている⁽⁹⁾⁽¹⁰⁾。しかし、実環境に近い環境モデルをシミュレータに構築するためには作業コストが高く、様々な環境で事前にシミュレーションすることは難しいと考えられる。そのため、河野らはロボットに搭載された複数種類のセンサを用いて環境形状や特徴情報などを取得してセマンティックマップを物理演算シミュレータ内に構築し強化学習するシステムを提案している⁽⁴⁾。河野らのシステムを用いることでヒトがシミュレータ内の仮想環境を作り込む必要が無く、ロボットが配置されている周囲の環境情報を自動的にシミュレータ内に構築可能となる手法である。さらにシミュレータによりロボットの移動経路等を強化学習し、自律移動後に再度環境情報を取得して次の移動経路シミュレーションを行う SMLO (sensing, mapping, learning, teleoperation) システムを実現している。河野らのシステムは複数種類のセンサをロボットに搭載しリッチな環境情報を取得しているが、LiDAR などの環境形状計測センサのみでも実行可能であり、本研究においても SMLO システムのフローを採用し、環境形状データの取得と強化学習による走行経路学習を行う。

〈2・3〉 本研究のアプローチ 従来研究や関連研究の課題を解決・活用し、本研究は次のような自律運転シニアカーシステムを提案する。シニアカーには LiDAR を搭載し、シニアカー周囲の 3 次元形状を取得する。取得した 3 次元形状情報から物理演算シミュレーション内に仮想環境を用意し、シニアカーモデルによる走行経路強化学習を行う。学習した知識を自律運転シニアカーに転移し、実環境での走

行を行う。学習した知識はサーバサイドに保存され、SAP-net を用いていつでも知識選択し自律運転シニアカーに転送できるように構成する。またサーバサイドで動作する SAP-net は従来の課題を解決するために並列計算を用いて計算量低減を行う。

3. 強化学習と転移学習

〈3・1〉 強化学習 本研究では学習アルゴリズムとして強化学習を用い、学習した知識(行動価値関数)を実シニアカーに転移するためには転移学習を用いる。これによりシミュレーションで学習したデータを実環境のシステムで活用する Sim-to-Real のフレームワークを実現する。強化学習は Q 学習を用い、次式で定義される。

$$Q(s, a) \leftarrow Q(s, a) + \alpha \{r + \gamma \max_{a' \in A} (Q(s', a') - Q(s, a))\} \dots (1)$$

ここで $Q(s, a)$ が強化学習で獲得する行動価値関数であり、全ての環境における行動とその価値が記録されている。 α は学習率、 γ は割引率、 r は報酬である。物理演算シミュレーションには Webots を用いる⁽¹¹⁾。Webots 内でシニアカーモデルが走行経路を学習し、獲得した行動価値関数を実シニアカーに転移する。

〈3・2〉 転移学習 転移学習は知識を再利用することで新たな問題の解決速度や適応力を高めるフレームワークである⁽¹²⁾。強化学習における転移学習は、あらかじめ Source task という環境でエージェントが強化学習した方策や行動価値関数を、新たな環境である Target task のエージェントが再利用する手法である。本研究では Taylor の転移学習を基本とし、次式のように行動価値関数の再利用を行う⁽¹³⁾。

$$Q_t(s, a) \leftarrow Q_t(s, a) + Q_s(s, a) \dots \dots \dots (2)$$

ここで $Q_t(s, a)$ は Target task すなわち実シニアカーが使用する行動価値関数で、新たに強化学習した行動価値は $Q_t(s, a)$ に記録される。 $Q_s(s, a)$ は転移される行動価値関数で Source task にて強化学習した行動価値関数である。すなわち Webots のシニアカーモデルが学習した行動価値関数である。

3. 提案システム

〈3・1〉 自律運転シニアカー 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の搭載を行うことで構築した。また、ET4D 本体には改造を加えずに自律運転機能を付与できるパッケージングとなっており新しい機能が付与されることを表すレトロフィット型となる(図1)。自律運転シニアカーには、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 3DLiDAR である Velodyne 社製 VLP-16 が搭載されている。また、転移学習時に必要となる学習環境における状態を表す自己位置座標と姿勢角の取得用として Intel 社製のトラッキングカメラである 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気社製測域センサ LRF UST-10LX が合わせて搭載されている。ハンドルの操舵にはアルミフレームから接続されたリニアアクチュエータを用いており、ワイヤ型のロータリーエンコーダと併用することでリニアアクチュエータの伸び縮みによる長さ制御を可能とし、シニアカー正面を 0 度とし-30 度から+30 度の範囲でのハンドル操舵が可能とされている。また、アクセルレバーの制御にはサーボモータを用いており、サーボモータに取り付けられたペダルを用いることでアクセルレバーを押し加減速が可能となっている。また実環境での実験を行うために、本研究で構築する計算機クラスタは著者らの研究室のサーバラックへ設置されているため、スマートフォンによるテザリング機能を用いて自律運転シニアカーに搭載されているコンピュータから VPN 接続で研究室内の計算機クラスタに接続する。

〈3・2〉 SAP-net の並列処理実装 これまで提案されている SAP-net は理論的には知識数に依存しないスケールな知識選択を実現可能であるが、シーケンシャルに知識を検索するような実装では計算時間の増大などの課題があった。そこで、本研究では SAP-net の計算処理を並列化し計算機クラスタに実装する。構築した計算機クラスタを図 2 に示す。計算機クラスタの構成として、Raspberry Pi3 model B+を 41 台搭載している。各 Raspberry Pi に搭載されている CPU の 1 コアを 1 ノードとするため、4 ノード×41 台で 164 ノード使用可能となる。また、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規格である MPI (message passing interface)を Python 言語で用いるため mpi4py を使用する。

〈3・3〉 全体システム構成 上述の自律運転シニアカーと計算機クラスタを用いて提案システムを構成するが、自律運転シニアカーに搭載された LiDAR から 3 次元環境形状データを取得し物理演算シミュレータでシニアカーモデルが強化学習するまでの流れを図 3 に示す。LiDAR で取得される出たはポイントクラウドと呼ばれる点群データであるため、Webots の環境モデルとして読み込めるよう処理される。その後、Webots 内の環境にシニアカーモデルを配置し、目的位置を設定してシニアカーモデルが実際に走行経路を学習する。

物理演算シミュレーションにて獲得された強化学習の行動価値関数は、LiDAR による環境形状計測時の LRF のデータの特徴データとして SAP-net の知識として保存される。実態としては計算機クラスタのノードの番号に対応付けられる形でサーバサイドに保存される。自律運転シニアカーは図 4 に示すように、実際の走行時に LRF のスキャンデータを計算機クラスタに転送し、計算機クラスは SAP-net 上で知識選択を実行する。計算機クラスタの SAP-net にて選択された知識はインターネットを介して自律運転シニアカーに転送される。自律運転シニアカーは転送されてきた知識を行動価値関数として転移学習し、実環境での走行を開始する。

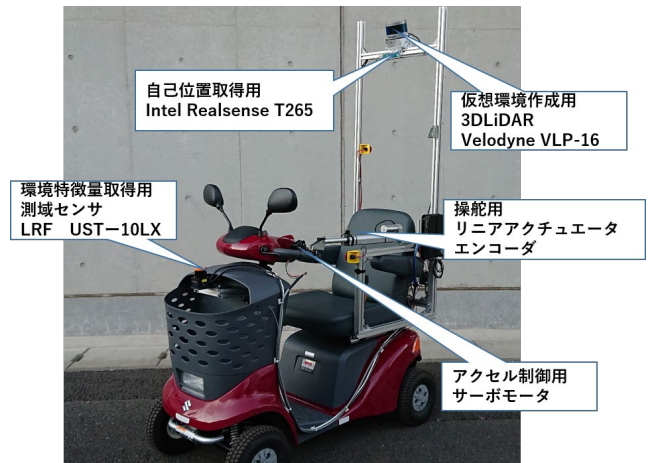


図 1 自律運転シニアカー



図 2 計算機クラスタ BitScope BR40

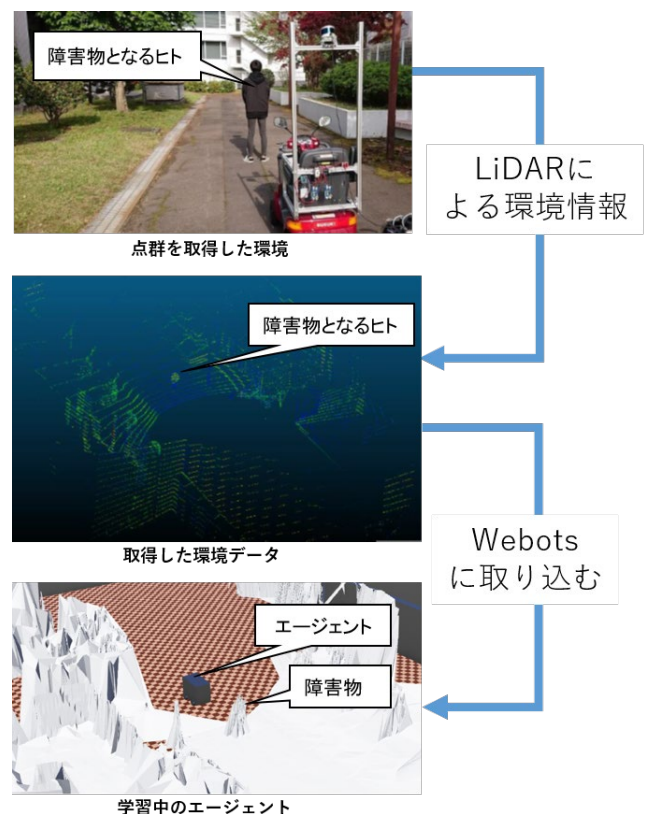


図 3 シミュレーションまでのフロー

4. 動作実験

〈4・1〉 **実験環境と実験条件** 提案手法の実環境における有効性の評価を目的に、自律運転シニアカーの走行実験を行った。実験条件として自律運転シニアカーに搭載された LiDAR により取得した環境データをもとに作成した物理演算シミュレータ内の仮想環境における強化学習を行い、得られた計 10 種の知識から知識選択を行うように SAP-net を構成した。また、知識選択における環境特徴量として自律運転シニアカーに搭載されている LRF から得られる環境形状データを用いる。実験環境として走行開始時には障害物等は存在しないものとし、走行開始後に障害物を配置する。これらの条件において環境特徴量から SAP-net が計算機クラスタ内で知識を選択し、選択された知識を用いて自律運転シニアカーが環境変化に応じて適切に知識を選択できるか評価を行った。

〈4・3〉 **実験結果と考察** 本実験における実験風景と移動軌跡を図 5 と図 6 にそれぞれ示す。結果として障害物の存在しない環境での走行中は、自律運転シニアカーが直進して走行する知識が選択された。走行途中に障害物を配し、知識選択において障害物を回避するための知識が選択され実際に障害物を回避しながら走行することが可能となった。これらのことから環境に対して有効な知識選択を行っていることが確認でき、提案手法が環境の変化に対して適切に知識選択を実行し自律運転シニアカーの環境適応性能向上に貢献していると考えられる。

5. おわりに

本研究では自律運転シニアカーの実現を目的に、シニアカーに搭載された LiDAR で計測した 3 次元環境形状情報から走行経路の強化学習を仮想的に物理演算シミュレータで実行し、学習結果である行動価値関数を実際のシニアカーに転移して自律走行を実現するシステムを構築した。さらにサーバサイドでは学習済みの知識が蓄積され選択される SAP-net を計算機クラスタによる並列計算実装をおこなった。これらのシステムを用いて評価実験を行い、走行中の経路上に歩行者などの物体が出現した時、歩行者を回避しながら走行する知識を自動で選択し環境に適応的な自律走行を実現するシステムであることが示唆された。

謝辞

本研究の一部は JSPS 科研費 JP19K12173 とスズキ財団令和元年度若手科学技術研究助成の助成を受けたものである。

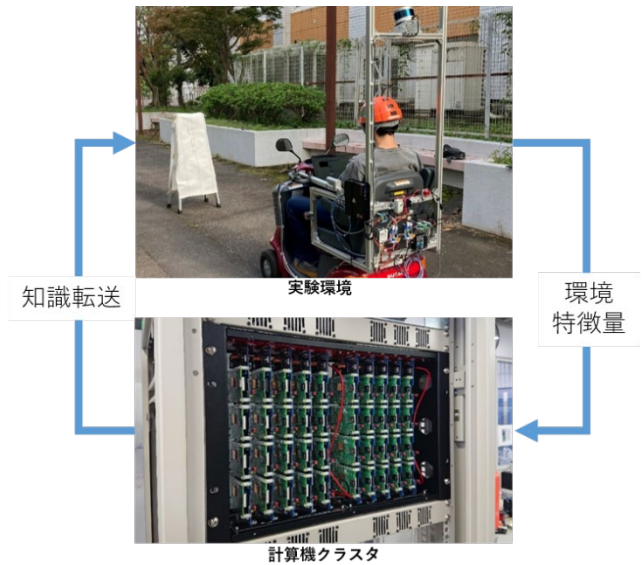


図 4 シニアカーと計算機クラスタの連携



図 5 実験風景

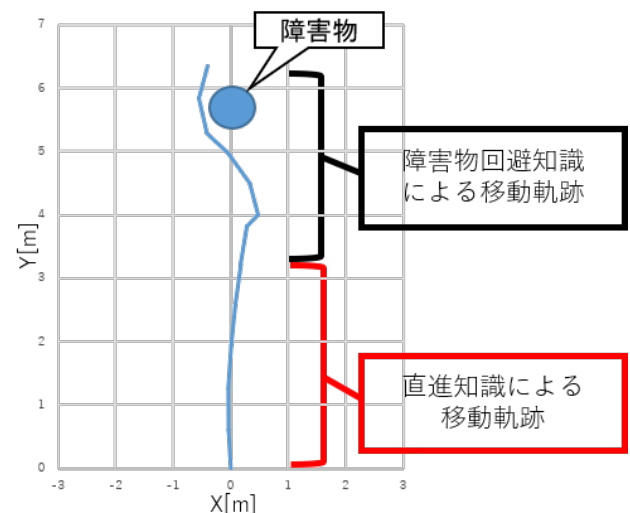


図 6 実験結果の移動軌跡

文 献

- (1) 浅田郁弥, 岡野憲, 中山雄一郎, 松日楽信人, 伊東敏夫, 廣瀬敏也, 「ロボットおよびシニアカーの連携システムの開発」, ロボティクス・メカトロニクス講演会, 2A1-F03 (2019)
- (2) 石田泰之, 伊東敏夫, 「全方位カメラを用いた走行環境認識のための信号機検出に関する研究」, 自動車技術会論文集, Vol.52, No.3, pp.695-700 (2021)
- (3) 沈舜聡, 伊東敏夫, 「センサーフュージョンにおける仮想濃淡勾配によるスパース点群の補間手法に関する研究」, 自動車技術会論文集, Vol.52, No.2, pp. 523-529 (2021)
- (4) 河野仁, 池勇勲, 藤井浩光, 「被災地情報収集のための半自律移動ロボットを用いたセマンティックサーベイマップ生成システムの開発」, 2021 年電気学会電子・情報・システム部門大会, pp.702-705 (2021)
- (5) F. Fernández, and M. Veloso: "Learning domain structure through probabilistic policy reuse in reinforcement learning", Progress in Artificial Intelligence, Vol.2, No.1, pp.13-27 (2013)
- (6) T. Takano, H. Takase, H. Kawanaka and S. Tsuruoka: "Effective Reuse Method for Transfer Learning in Actor-critic", SCIS & ISIS, pp.137-141 (2010)
- (7) 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, Vol.10, No.12, pp.7-16 (2019)
- (8) Y. Takakuwa, H. Kono, H. Fujii, W. Wen, and T. Suzuki: "Autonomous Reusing Policy Selection Using Spreading Activation Model in Deep Reinforcement Learning", International Journal of Advanced Computer Science and Applications, Vol.12, No.4, pp.8-15 (2021)
- (9) J. v. Baar, A. Sullivan, R. Cordorel, D. Jha, D. Romeres, and D. Nikovski: "Sim-to-Real Transfer Learning using Robustified Controllers in Robotic Tasks involving Complex Dynamics", 2019 International Conference on Robotics and Automation (ICRA), pp.6001-6007 (2019)
- (10) M. Hwasser, D. Kragic, and R. Antonova: "Variational Auto-Regularized Alignment for Sim-to-Real Control", 2020 IEEE International Conference on Robotics and Automation (ICRA), pp.2732-2738 (2020)
- (11) Webots: robot simulator, <https://cyberbotics.com/> (June,15,2022 Access)
- (12) Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, Transfer Learning, Cambridge University Press (2020)
- (13) M. E. Taylor: "Transfer in Reinforcement Learning Domains", Springer, Vol. 216 (2009)
- (14) 坂本裕都, 河野仁, 温文, 藤井浩光, 鈴木剛, 「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」, 2021 年電気学会電子・情報・システム部門大会, pp.711-716 (2021)