

能動的・受動的エージェントとのすれ違いにおける 被験者の行動戦略の分析

Analysis of subjects' behavioral strategies in collision avoidance with active and passive agents

宮本賢良^{1*} 渡邊紀文²
Kensuke Miyamoto¹ Norifumi Watanabe²

¹ 慶應義塾大学

¹ Keio University

² 武蔵野大学

² Musashino University

Abstract: 人が他者と協調する際には、相手の様子を確認しながら行動する受動的戦略だけでなく、自らの行動方針を示すために先に行動する能動的戦略がある。例えば道ですれ違う際には、相手を避けるだけではなく、能動的に先に左右へ移動する行動が考えられる。本研究では、コンピュータが操作する受動的・能動的、2種類のエージェントと、被験者があい路ですれ違うシミュレーション実験を行った。人が他者と協調する際の行動パターンを分類しモデル化するため、コンピュータエージェントの戦略に応じた、被験者の行動戦略を分析した。

1 はじめに

ロボット技術が発展することで、近い将来、人々の生活に密着してサポートするようなロボットも登場すると予想される。生活環境で用いられるロボットは、人と協調して作業を行う機会も増加すると考えられる。

本研究では、道ですれ違う際の行動戦略について考えている。そのため、コンピュータが操作する受動的・能動的、2種類のエージェントと、被験者があい路ですれ違うシミュレーション実験を行い、以前の分析では、同じ対面状態でも異なる行動選択をする被験者とならない被験者がいることがわかった [1]。本発表では、場面に応じた被験者の行動をより細かく分析することで、人の戦略がインタラクションを通じて相手の振る舞いに適したものに变化していることを確認する。

2 背景

人同士の協調行動では、インタラクションをする時には、お互いに相手の行動に従って反応する。相手に合わせる受動的な行動戦略だけでなく、自らが先に行動する能動的な行動戦略があり、人はそれらの戦略を切り替えて、時には他者を自らの目的に合わせて振る

舞うように誘導するといった協調行動も取っている [2]。人と生活の場を共にするロボットも、人のように複数の戦略を切り替え、相手の戦略にかみ合わせ、競合を避ける行動戦略が適していると考えられる。筆者らが取り組んだ以前の研究では、人がすれ違い時に、相手の戦略に応じた行動戦略を取っていることは確認できた [1]。しかし、エージェントとの協調戦略を獲得する過程については不明なままである。

メタ戦略は、表面的な行動決定過程の背後にある戦略であり、人はこのメタ戦略に基づいて戦略を決め、その戦略によって行動を決定している。メタ戦略モデル [2] には受動的戦略と能動的戦略が定義されている。受動的戦略は観察をもとに他者の意図を推定し、それを考慮して自身の意図を決定、それを達成するための行動を取る。しかし、協調戦略の切り替え方となるメタ戦略自体や、ベースになる知識はどのような環境で学習すべきか・相手の立ち位置で特定の行動を取った時に環境がどう変化するか、さらにその場面で適切となる行動をどのように学ぶかなど課題が残っている。

3 能動的・受動的エージェント

本研究ではすれ違いを協調行動のタスクの例として扱う。受動的なすれ違い戦略は、相手が避ける方向を確認し、ぶつからないように自らの移動方法を定める

*連絡先：慶應義塾大学
神奈川県藤沢市遠藤 5322
E-mail:kensuke-m@keio.jp

ことが考えられる。一方で能動的な戦略の時は、先に道の片側に寄ることによって、どちらを通るか伝えることでなめらかなすれ違いが可能になる。さらに、他者の目的や、他者が取ると予想される行動などを考慮しない戦略も、1つの協調時の戦略として考慮する。行動戦略を、他者と上手にかみ合わせることで衝突などの不都合を回避することが出来る。しかしながら、どのような目的を持っているか、どのような行動戦略に従っているかは、口頭などで明示的に与えられる場面は少なく、相手の振る舞いから推定する必要がある。

人との協調実験において対向者となる、コンピュータが操作するエージェントの行動戦略は、Q学習 [3] を用いて、あらかじめコンピュータ同士で学習を行った。メタ戦略モデルの考え方を取り入れ、まず単純に時計回り・反時計回り両エージェントが学習する試行を行い、そこで得られた戦略を相手が取ると想定した環境で単体のエージェントだけが学習することで、単純な振る舞いをする相手の行動の影響を踏まえた上での行動戦略の学習を行った。さらにもう一段、その戦略に対する協調戦略を学習させることで、相手に道を譲る戦略と、譲らせようとする行動戦略を学習した。

4 人のすれ違い戦略行動実験

4.1 実験目的

本研究では、人の場合は自らの経験を元に他者とすれ違う戦略をすでに獲得済みでありつつ、それを元に相手に合わせた調整を行っているという仮説を立てている。そこで、人がエージェントとインタラクションをする過程で、複数の戦略をどのように切り替えていくのかを調べるために、マス目状の回廊を周回する環境において、人が逆方向へ向かうエージェントと対面した時に、どのような行動選択をするかを調べる実験を行った。

4.2 実験内容

本研究では、マス目状の回廊を周回する環境 (図 1) において、人が逆方向へ向かうエージェントと対面した時に、どのような行動選択をするかを調べる実験を行った。様々な位置関係での被験者の行動選択を計測するため、被験者とエージェントの配置は 10 回の行動選択 (ステップ) で再配置される。この実験で主に使用するのは一辺のみである。再配置直後の配置パターンは図 1 の 1 から 5 のいずれかに被験者エージェントが配置されることによって変わるあい路までの位置関係と、コンピュータが操作するエージェントが配置される距

離 (隣接・1 マス離れる) の 2 通りを掛け合わせた 10 通りである。向きは常に対面状態で配置される。

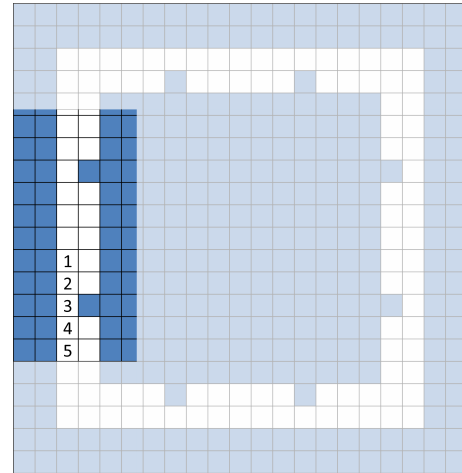


図 1: 実験環境. フォーカスした左辺の部分ですれ違う。1 から 5 のいずれかに被験者エージェントが配置される。実験ではあい路がある経路を内周、あい路がない経路を外周として説明している。

エージェントは自らの正面 2 マス、左右と後方 1 マスの周囲 3×4 マスの状況を観測することができる。区別できる各マスの状態は、何もない、壁、時計回りエージェント、反時計回りエージェントの 4 種類である。エージェントには上下左右の向きがあるが、視界内の他のエージェントがそれぞれどちらを向いているかは観測できない。エージェントは 1 マス前または後ろに移動、左右へ 90° 方向転換、停止のいずれかの行動を選択する。移動を選択した場合は何もないマスにのみ移動でき、移動先に壁や他のエージェントがいた場合は移動は行われない。ステップごとのエージェントの行動決定は、すべてのエージェントが行動する前にを行い、ステップ内の行動順はランダムに決定される。

被験者は図 2 のようなインターフェースを持つアプリケーションを用いて時計回りに周回するエージェントを操作する。実験用のアプリケーション (図 2) の左部は、被験者が操作するエージェントの周囲の状態を表す。回廊の壁は青の四角、被験者の操作するエージェントは黒丸、コンピュータが操作するエージェントは白丸で表す。例えば図 2 の状態は、被験者エージェントが図 1 の辺上のあい路を通り抜けた地点で、コンピュータエージェントと隣接して対面している状態を表している。対向者となるコンピュータが操作するエージェントには、進むべき方向の判断材料として内周側と外周側どちらにいるかという情報が与えられるが、被験者には立ち位置の判断を容易にするため、周回する時にどの方向に向かえば良いかという情報を、画面右部

に矢印で示した。

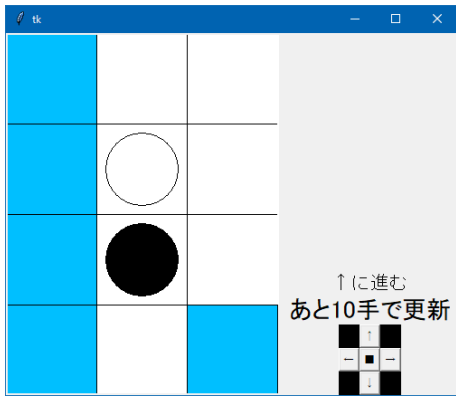


図 2: 被験者がエージェントを操作するインターフェース

4.3 被験者の行動選択分析結果

以前の分析 [1] では同じ対面状況でも被験者が様々な行動を取っていることがわかった。そのことから、コンピュータエージェントの取る戦略に応じた行動選択を行っていることが示唆された。そこで本発表では本当に被験者がコンピュータエージェントの戦略に基づいて行動を変化させているのかを調べるための行動選択を分析した。図 3, 4 は、同じ状態に対して何種類の行動が選択されているか、その割合を示す。各被験者に対して、前後半 300 回ごとに分けて集計を行った。そもそも発生回数が少ない (3 回未満) 状態は、選択される行動が前進後退・左右旋回・停止のうちの 1 つになりやすいため、集計から除外した。また、行動の選択回数に偏りがある場合も見られたが、選択回数が極端に少ない (2 割未満) 行動については選択された行動とは扱わない。図 3 は被験者とコンピュータそれぞれが操作するエージェントの配置に基づいた状態ごとに集計した。図 4 は、コンピュータエージェントの戦略が異なる場合は、別の状態として扱い集計を行った。

表 1 は、実験の前半後半それぞれ 300 回での、2 ステップ目での行動選択までの平均時間である。言い換えれば、再配置後に 1 度行動選択し、その次の行動選択までの時間である。再配置のタイミングでコンピュータエージェントの戦略が変わる可能性がある。その瞬間には被験者がコンピュータの戦略の変化を認識することは出来ないが、1 度両者が行動した場面で予想していた行動と異なる行動を相手が取った場合に、相手の戦略の推定値を修正する必要がある。その場合判断に時間がかかると考えた。ただし、行動選択までの時間が人の反応速度と言われる 0.2 秒未満の場合と、3 秒以上の場合を外れ値として集計からは除外した。表 2 は

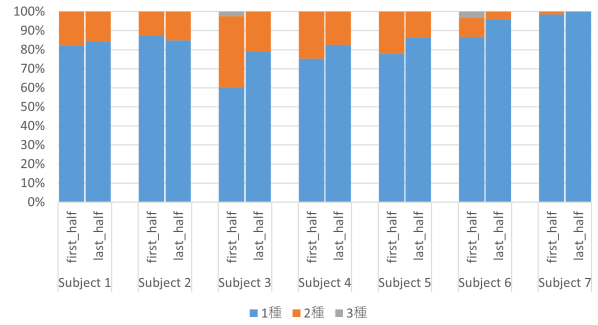


図 3: エージェントの配置に基づいた状態ごとに選択された行動の種類の数

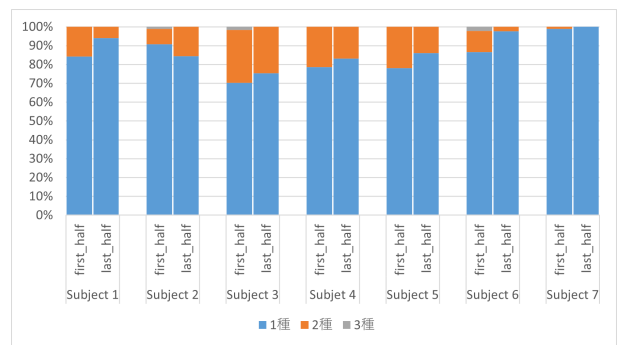


図 4: コンピュータエージェントの戦略も状態に加えた場合の選択された行動の種類の数

初期配置別の 2 ステップ目での行動選択までの平均時間である。表内の 3, 4 は図 1 の配置箇所を表し、あい路においてエージェントが隣接しているもしくは 1 マス空いている状態を示す。

4.4 考察

図 4 から、被験者 1 は実験後半では、コンピュータエージェントの取っている行動戦略を含めた状態とすることで、ほとんどの場面での行動選択が 1 種となっていることがわかる。周囲の状況のみを状態の区別に使用した場合 (図 3) では 2 割程度の状態で行動選択が 2 種となっている。このことから被験者 1 は、コンピュータエージェントの行動戦略をその行動から読み取り、予測した相手の戦略によって自らの行動選択も変化させていることがわかる。

被験者 2, 3, 4, 5 は前半だけでなく、後半においても、コンピュータエージェントの戦略を考慮した場合でも 2 種類以上の行動選択をする状態が残っている。このことから、特に今回の実験環境でのコンピュータエージェントと協調するための戦略の学習が、まだ進行中

表 1: 実験前半・後半別のステップ 2 の行動選択までの平均時間

	first_half	last_half	all
Subject 1	0.264	0.245	0.255
Subject 2	0.437	0.553	0.495
Subject 3	0.680	0.481	0.580
Subject 4	0.531	0.465	0.498
Subject 5	0.321	0.311	0.316
Subject 6	0.510	0.463	0.486
Subject 7	0.382	0.359	0.370

表 2: 初期配置別, ステップ 2 の行動選択までの平均時間. 3, 4 は図 1 の各所.

	隣接配置		1マス空けて配置	
	3	4	3	4
Subject 1	0.273	0.223	0.294	0.265
Subject 2	0.680	0.606	0.479	0.521
Subject 3	0.469	0.582	0.684	0.542
Subject 4	0.480	0.419	0.628	0.448
Subject 5	0.325	0.103	0.610	0.547
Subject 6	0.319	0.337	0.640	0.717
Subject 7	0.436	0.282	0.702	0.365
average	0.426	0.365	0.577	0.486

で十分な戦略が形成されていない可能性が高い。被験者 3, 4, 5 は後半の方が行動選択が複数種類になっている状態が少なくなっているため、学習は進んでいると考えられる。特に、被験者 3 は前半と後半の行動選択までの時間が短くなっていることから、コンピュータへの対応を学んでいると考えられる (表 1)。

一方で被験者 2 のように、後半の方が選択までの時間が長くなっている例もある。個々の開始状態について確認すると前半に比べて後半の選択時間が長くなっている状態が複数見られた。インタラクションを通して、コンピュータエージェントに特に見られる振る舞いを学んだことで、むしろ選択する行動を悩む状態があると考えられる。

被験者 6 は後半になってから、被験者 7 は前半も含め実験の開始時点から、周囲の状態ごとに行動選択が 1 種類となっている。このことからコンピュータエージェントの戦略は、被験者行動選択には反映されていないことがわかる。ただし、今回の実験では、被験者が道を譲ろうとするのと同じタイミングで、コンピュータも道を譲った場合、しばらく道を譲り続けることが明白になる。コンピュータは「被験者が譲ってくれるならば、自分は譲ってもらう戦略に変更する」のような、その場その場に応じた戦略レベルでの切り替えは行わ

ない。一概に受動的な振る舞いをする戦略と表現しても、相手の行動に対して違う経路を取る受動性や、相手の戦略に対してそれに応じた戦略を取る受動性など、様々な受動性が考えられる。今後、それらの戦略自体を切り替えるモデルを学習する方法や、そのモデルに基づいたエージェントの評価方法についても検討する必要がある。

状態 3, 4 における選択時間については、コンピュータエージェントの配置によって比較すると、被験者 2 以外は隣接配置の方が時間が短くなっている (表 2)。これは、隣接配置ではコンピュータがあい路を超えるのを被験者が邪魔をしている場面なので、自らが道を譲るのが自然な場面であるためと考えられる。一方、距離がある配置の時は、どちらが道を譲るかは、互いの戦略に依存しているため、戦略を判断する分行動選択には時間がかかる。状態ごとに戦略の選択に影響が大きいと考えられる。

5 まとめ

本研究では、人が他者と協調する際の行動戦略をどのように獲得しているかを調べるため、実験を行った。分析結果から、コンピュータエージェントの戦略に応じて行動選択が一貫となり、相手の戦略の推定値に基づいて自らの行動を変化させていることが示唆された。しかし、インタラクションを多く済ませ学習が進んだと考えられる場面でも行動選択にばらつきが出たり、状態によって依存する戦略の違いにより、行動選択の時間に差がある被験者も見られた。今後は、戦略自体を切り替えるモデルの学習過程や評価方法や、状態ごとにどのような戦略が選択されやすいかについても検討する。

参考文献

- [1] Kensuke Miyamoto, Norifumi Watanabe, Osamu Nakamura, and Yoshiyasu Takefuji. Analysis of a human meta-strategy for agents with active and passive strategies. *Applied Sciences*, Vol. 12, No. 17, p. 8720, 2022.
- [2] Ayami Yokoyama and Takashi Omori. Modeling of human intention estimation process in social interaction scene. pp. 1-6, 07 2010.
- [3] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.