

TSOMを用いた行動戦略解析

九州工業大学大学院
生命体工学研究科 生命体工学専攻
博士論文

指導教員: 石井和男

富永萌子

2021年3月31日

© 2020

富永萌子

審査委員会

主査: 石井和男

副査: 我妻先生

副査: 林英治先生

謝辞

本研究を遂行するにあたり，熱心なご指導と有益なご助言を賜りますとともに，研究者としての心得を御教授してくださいました九州工業大学大学院生命体工学研究科人間知能システム工学専攻 石井和男教授に心から感謝します。

目次

謝辞	iv
第1章 序論	2
1.1 はじめに	2
1.1.1 ロボットの社会実装	2
1.1.2 協調行動	3
1.1.3 ロボット学習	3
1.1.4 人間ロボット協調	4
1.1.5 ロボカップ	5
1.2 本研究の目的	6
1.3 本論文の構成	6
第2章 人間のチーム行動解析	7
2.1 自己組織化マップ (SOM)	7
2.1.1 競合学習	8
2.1.2 アルゴリズム	8
第3章 人間と同じ行動決定の創発	10
第4章 ロボットへの実装実験	11
第5章 考察及びまとめ	12

Abstract

With the progress of technology, the realization of a symbiotic society with human beings and robots sharing the same environment has become an important subject. An example of this kind of systems is soccer game. Soccer is a multi-agent game that requires strategies by taking into account each member's position and actions. In this paper, we discuss the results of the development of a learning system that uses SOM to select behaviors depending on the situation. This system can reproduce the action selection algorithm of all players in a certain team, and the robot can instantly select the next cooperative action from the information obtained during the game. Because of this system, common sense rules were shared to learn an action selection algorithm for a set of agents, not only a team consisting of robots alone, but also a group of heterogeneous agents consisting of humans and robots.

第1章 序論

1.1 はじめに

1.1.1 ロボットの社会実装

近年、安全安心で持続可能な社会の実現、少子高齢化対応や第一次産業を始めとした産業基盤の構築等の社会的な課題に対する解決策として、ロボットの社会実装が期待されている。ここで、ロボットの社会実装の社会的期待について考えるため、工場内でのロボット有用を例にとる。産業用ロボットの普及に伴い、工場内でのロボット実装には労働者を守るための安全対策が複数関わってくる。代表的なものとして、労働安全規則や国際標準化機構 (International Organization for Standardization: ISO)、日本工業規格 (Japanese Industrial Standards: JIS) などが挙げられる。労働安全衛生規則第 150 条の 4 では、
事業者は、産業用ロボットを運転する場合（教示等のために産業用ロボットを運転する場合及び産業用ロボットの運転中に次条に規定する作業を行を行わなければならない場合において産業用ロボットを運転するときを除く。）において、当該産業用ロボットに接触することにより労働者に危険が生ずる恐れのあるときは、さく又は囲いを設ける等当該危険を防止するために必要な措置を講じなければならない。）
とある。この規則により、人間とロボと slack が同じ工場内で作業する場合において、その間には、柵や囲いを設ける必要があり、協働することは困難であった。しかし、規制改革実施計画及び近年の技術改心が加味され、平成 25 年 12 月 24 日付基発 1224 第 2 号通達によりいかが明記された。

産業用ロボットを使用する事業者が、労働安全衛生第 28 条の 2 による危険性等の調査（以下「リスクアセスメント」という。）に基づく措置を実施し、産業用ロボットに接触することにより労働者の危険の生ずるおそれが無くなったと評価できるときは、本条の「労働者に危険が生ずるおそれがあるとき」に該当しないものとする。（以下省略）

したがって、現在では安全が確保できていると確認できる環境下であれば、人間とロボットが協働作業可能となった。

このように、法整備という観点から観察しても、社会的期待としては人間とロボットが同じフィールドで安心して作業や生活さえも行える将来が望まれている。しかし、度ボットにか

んする研究開発では、「人とロボットの関係性」についてどこまで考えられているであろうか。ロボットの社会実装には、社会的期待と研究開発との方向性の乖離を避け、研究成果をわかりやすく社会に還元・提示する必要がある。そのため、人間とロボットをどのように共存させるか、人間ロボット共生社会のあり方を議論すべきである。

1.1.2 協調行動

協調行動は、異なる自律エージェントが共通のタスクを実行しながらコミュニケーションを図るときに重要な側面である。多くの場合、単一のエージェントでのタスクを実行は必ずしも効率的とは言えず、ここ数年で、困難な問題を解決するためにマルチエージェントシステム(MAS)を研究している研究者も多い。マルチエージェントシステムは、複数のエージェント(自律エージェント)が協調行動によって共通の目標を達成しようとするシステムである。センサーから取得したデータに基づいてリアルタイムで決定を下すことによって、エージェント同士に加え、環境と相互作用する。(?) (?) MAS のテストベッドとして有名なRoboCupは、世界的ロボット大会であり、強化学習やニューラルネットワークなどの学習方法を使用してマルチエージェントの協調を一つの課題としている。RoboCupはランドマークプロジェクトであり、2050年までに、人間のサッカーワールドカップチャンピオンチームに勝つロボットチームの制作プロジェクトである(?)。例としてこれまでに、Sandholm and Crites(?)によれば、十分な測定データと行動が利用可能であれば、強化学習は反復囚人のジレンマに対して最適な手法であると示された。さらに、Arai(?)は、環境がグリッドとしてモデル化されている場合のマルチエージェントシステムにおける追跡問題に関するQラーニングと利益共有の方法を比較し、協力的な行動が利益共有の間で明確に現れることを示した。しかしながら、これらの研究は、実環境で動作するロボットへの応用をまだ考慮されていない。

1.1.3 ロボット学習

学習アルゴリズムの中で、教師なし学習はMASシステムにとって有望な方法である。教師なし学習の利点は、ロボットが環境に関する以前の情報やロボット自体の前提知識を持っている必要がないことである。しかし、学習には多くのテストや経験が必要であるため、設計者は、状態空間とアクション空間を定義するために、試合中に考慮すべき重要なパラメータと変数を明確に定義することがコストと時間を削減するために重要である。状態空間の設計は、ロボットが行動空間で何ができるかによって異なる。同時に、状態空間はロボットの能力とそれ自身の行動空間に従って定義される。2つの空間は互いに相互接続している。効率的な方法で状態空間を設計するために、Asada(?)は、状態空間を最初に2つの主要な状態に分割し、次に状態数

を増やししながら再帰的に多数の層に分割する方法を提案した。しかし、いくらかの偏りの問題は依然としてロボットの行動に影響を及ぼし、何らかの誤った行動および習慣を引き起こす可能性があると示された。

1.1.4 人間ロボット協調

上述したように、人間 - ロボット共生社会は多くの研究者にとって関心のあるトピックであり、いくつかの研究が行われてきた。したがって、そのような共生システムでは、ロボットが人間の行動を理解し解釈し、それに応じて行動することが必須である。しかし、これまでの研究では目標として、ロボットの行動のみ、または特定の数の人間とロボットの間で正確に定義されたインタラクションに基づく行動に関することが大半であった。また、マルチエージェントシステムでは、人間とロボットとの間の通信は、通常、ある種のインタフェース、例えばコマンド音声またはジェスチャによって行われている。しかし、システムが非常に複雑な場合は特に、これでは不十分な場合が存在する。したがって、次のステップとしてロボットに、人間がそうであるように、いくつかの状況を理解し、その挙動に予測的に適応させることを目指す。この研究の主な目的は、人間同士の協調行動を考え、それをロボットに应用することである。

1

ニューラルネットワークへの入力ベクトルの要素は、人間とロボットの行動との間の可能な最小のギャップを達成するように選択される。そのためには、まず人間の行動を理解し、そのような行動を理解して模倣できるロボットを開発することが重要である。文献(?)では、フットサルゲームにおける人間とロボットの行動の研究に焦点を当てている。なぜなら、これは動的な環境、いくつかの制約を伴う優れたテストベッドであり、リアルタイムの計画が必要だからである。これらは、ロボットが将来動作する可能性がある一般的な共生システムの特性となる。使用されるアルゴリズムでは、スコア、コーナーキック、ペナルティキック、ボールを持っているチームなど、ゲーム内の特定のイベントに応じて、SOMを使用してプレイヤーのポジションが分析された。その結果、人間の試合もロボットの試合も同様に、試合における場面を評価することが可能であり、学習する入力要素を増やすことによって、環境と行動の写像関係を示すSOMの作成によってロボットによる自律的行動の発現が可能であることが判明した。この論文では、その写像関係を求め、テンソル自己組織化マップによりロボット

¹補足があればここに...

同士の試合を学習し、サッカーロボットの行動学習システムの開発を行う。

1.1.5 ロボカップ

ロボカップとは「2050年までに人間のW杯優勝チームにかつ自律ロボットチームを作る」という大きな目標を掲げた国際ランドマークプロジェクトである。サッカーというテーマは、動的な環境下で囚人に判断をする知能や、マクドナルドシステムによるチーム行動といった要素を有している。しかし、サッカーというテーマだけでなく、ロボカップでは、人間のせい買空間であるキッチンやリビングを再現したフィールドで競技を行う「@ホームリーグ」や、災害現場で教授に役立つ自律ロボット開発を推進する「レスキューリーグ」、次世代の人材育成を目的とした「ジュニアリーグ」、2012年から追加された「インダスリアルリーグ」など、「サッカーリーグ」を含めて現在では5つのリーグが存在し、様々な分野の技術を発展させるプロジェクトとなっている。各リーグで共通する課題は、「複数ロボットの協調によって目的は達成され、かんきょうが動的に変化し、多目的/多重制約問題を包含し、実時間計画/推論を必要とすること」である。

筆者が所属するチーム「Hibikino-Musashi」は、サッカーリーグの中の、「ロボカップサッカー中型ロボットリーグ (Middle Size League)」(Fig.) である。

ロボカップサッカー中型ロボットリーグ (MSL)

中型ロボットリーグ（以下、MSLと表記する）では、年々フィールドのサイズが大きくなっていくが、現状では12[m] × 18[m]の大きさのフィールドで試合を行うリーグである。各チームの試合に参加可能なロボットは5台までとなっており、ロボットの大きさは52[cm] × 52[cm]、高さ80[cm]以下となっている。（ゴールキーパーロボットは1秒間のみ、60[cm] × 60[cm]、高さ90[cm]まで拡大可能）ロボットは搭載されたカメラなどのセンサ情報から環境認識し、自律行動を行う。試合は前後半15分の合計30分間行われる。

試合データ

ロボカップサッカー中型リーグでは、2017年から試合中のロボットの自己位置・障害物認識状態などの情報をリアルタイムに近い状態で、提出することが義務かされるようになった。これは、試合中のデバッグ用だけでなく統計分析のためのデータ標準化である。提出された全参加チームデータは、インターネット上に公開され、研究・開発その他に誰でも利用可能である。ロボットや障害物、ボールに関する全ての座標データは、どのチームも定義された3次元空間のデカルト座標系で提出しなければならない。

1.2 本研究の目的

本研究の目的は、人間のチームとしての行動決定アルゴリズムを解析し、教師なし学習である自己組織化マップを用いて、近似されたモデルを用い、ロボットに環境人氏から行動決定の写像関係を学習させることを目的とする。

1.3 本論文の構成

本論文の構成は、第2章では、本研究において使用する自己組織化マップについての歴史とアルゴリズムを述べた上で人間のチーム行動の解析を行う。第3章では、第4章では、最後に第5章では、考察及びまとめを述べる。

第2章 人間のチーム行動解析

2.1 自己組織化マップ (SOM)

自己組織化とは、局所的な相互作用から地域的に順序づけられた構造が導かれることであり、複数の要素から構成されるシステムが他からの制御を得ることなく、時間とともに何らかの意味で自発的に秩序化する過程である。神経回路の構造は、脳内においてこの自己組織化の代表的な例として挙げられる。1956年に Sperry[s] によってはじめられた視覚神経通路の発達における位相的順序を保持した状態での網膜位相マップの構築を目的としたモデル化が、自己組織化の位相形成原理についてのモデル化の先駆けとなり、Von der Malsburg[?] が特徴選択における皮質細胞の局所的な順序づけを目的としたシナプス学習を含む自己組織化過程を導入するなど、創刊がある神経活動によって駆動される位相的なマップの形成に関係した自己組織化の要素が神経活動による位相マップの初期の形成には存在するという結論が主流となった。その後、1976年に Willshaw と von der Malsburg[?] によって、本質的には同様の幾何学的関係がある離散的な格子間での類似関係を検出する自己組織化処理モデルが、自己組織化マップ (Self-Organizing Map : SOM) として提案され、SOM は、教師なし学習であり、生命の神経回路による情報処理に近い競合学習として知られるようになった。

Kohonen は前述したモデルからさらに、類似関係があるものではない、かつ幾何学条性格である必要もない離散的格子上に連続した入力君間からマップ化することを目的として学習アルゴリズムを変更した [?]. より強度な自己組織化効果を得るために、近傍関係を学習に導入するため荷重更新規則を修正し、これらの近傍関係によって広がった範囲が時間経過にともない縮まる要素も追加した。これにより、SOM のアルゴリズムとして、Kohonen モデルが一般的となり、本論文中にあいても Kohonen モデルを取り扱う。

この構造により、多次元入力ベクトルの低次元化が出力そうにて可能であり、特徴が近似している入力ベクトルほど、近くに配置される。SOM の応用例としては、複雑なデータの二次元表示への可視化があり、クラスタリング技術と同様に、抽象化の創造が可能である。優れた補間性能を保つために、学習データの一般化が可能であり、未知な入力データに対しても妥当性の高い出力が期待される。

2.1.1 競合学習

競合学習は階層型ニューラルネットワークにおいて重要な懸念である。同一の入力に対し、各々のニューロンが活性度を競い合う側方相互作用が利用される。具体的に競合学習としての SOM は、入力層と出力層からなり、入力層 (入力空間 V) の各入力サンプル (入力ベクトル $v(t)$) は全ての出力層ニューロンと結合し、結合には結合荷重 w_i が付加されている。ここで t は時間座標であり、 N 個の出力層ニューロンは、 $i=1, 2, \dots, N$ とラベル付けされている。(Fig 参照) そして、近くのニューロンとの結合は興奮性 (正の結合荷重) があり、より遠方のニューロンとの結合には抑制性 (負の結合荷重) がある状態で出力層のニューロンは活性化するために競合する。

2.1.2 アルゴリズム

Kohonen の SOM アルゴリズムは、大きく競合段階と協調段階の二つの改装で構成される。最初の段階ではまず、入力に対して、最整合ニューロン (Best Matching Unit : BMU) の選択すなわち勝者を選択する。次に、その最整合ニューロンおよびその直近の格子の荷重変更が行われる。以下に、これら二つの段階について述べるが、本研究では、新しい入力ベクトルの投入のたびに、荷重更新を行う増分学習 (オンライン学習) ではなく、全入力ベクトルの協業段階が行われた後に、生成できる全入力 that 考慮された荷重更を行う、バッチ型 SOM [?] を用いる。これにより、入力ベクトルの投入順番による出力への影響を考慮する必要がなくなる。よって、以降は本研究で用いたバッチ型 SOM について述べているため、バッチ型 SOM への変更に伴い、入力空間 V を $\mu = 1, \dots, M$ の M 個のサンプルからなる固定の学習ベクトル集合 $\{v^\mu\}$ として定義する。

競合段階

競合段階では、各入力ベクトル v^μ に対する最整合ニューロンについて式 (1) を用いて選択する。BMU の決定方法としては、内積距離規則やユークリッド距離規則が考えられる。内積距離規則は生物学的モデルに類似しているため、多くその方面で用いられているが、今回は生物学的なモデルとして SOM を扱う必要がないため、ユークリッド距離規則による BMU 決定を採用する。ユークリッド距離を用いる場合、結合荷重改め、ベクトル量子化の観点から参照ベクトルと呼ぶこととする。ここでは、勝者 j^* とラベル付けする。また、 $\|\cdot\|$ はユークリッド距離を表している。

$$i^* = \operatorname{argmin} \|w_i - v^\mu\| \quad (1)$$

協調段階

協調段階では、ニューロンが発火 (BMU が決定されること) することにより、その周囲のニューロンも協調的に発火しやすくなるように荷重更新に影響を与える段階である。各丹生論の参照ベクトル w_i を式 (2) により更新する。ここで、 t は学習回数を示す変数であり、総学習回数を t_{max} とする。 η は学習係数である。

$$w_i(t+1) = w_i(t) + \frac{\eta}{M} \cdot \Delta w_i \quad (2)$$

ただし、参照ベクトルの増減値 Δw_i は、次式 (3) で示される。

$$\Delta w_i(t+1) = \Delta w_i(t) + \Lambda(i, i^*, \sigma_\Lambda) \cdot (v^\mu - w_i) \quad (3)$$

ここで、 Λ はニューロン i と i^* の格子座標 r と r^* での尺度調整を行う近傍関数である。勝者の位置に関して左右対称であり、勝者からの格子距離が増加するにしたがって単調に減少する関数でなければならない。今回は、もっとも一般的なガウス型近傍関数 (式 4 参照) を用いた。(図 [?] 参照) σ_Λ は、その範囲 (標準偏差) である。

$$\Lambda(i, i^*) = \exp\left(-\frac{\|r_i - r_{i^*}\|^2}{2\sigma_\Lambda^2}\right) \quad (4)$$

先に述べた、近傍関数によって広がった範囲が時間経過にともない縮まる要素として、標準偏差、式 (5) によって学習回数とともに小さな値をとる。(図?参照)

$$\sigma_\Lambda = \sigma_{\Lambda_0} \exp\left(-2\sigma_{\Lambda_0} \frac{t}{t_{max}}\right) \quad (5)$$

第3章 人間と同じ行動決定の創発

人間と同じ行動決定創発を書き書きします

第4章 ロボットへの実装実験

TSOM を用いて行った解析をいかに示す

第5章 考察及びまとめ

まとめるぜ !!!!