

論文の内容の要旨

論文題目 実際のコーチング法に基づいたサッカーエージェントによる
協調戦術の獲得

氏名 熊田 陽一郎

本研究の目的は、サッカーにおける協調戦術の学習のモデル化を通して、マルチエージェントシステムにおける学習による協調作業の獲得機構、および、協調学習の新しい方法を提案し、評価することである。

サッカーエージェント研究は、協調行動を特徴とするサッカーという課題をロボットやエージェントにより遂行させることを目的とする。それゆえ、その研究のアプローチは、マルチエージェントによる協調学習の問題に対する貢献が当初は期待された。しかしながら、現在のサッカーエージェント研究における、学習の目標は、主にフォーメーションと役割（いわゆるフォワード、ディフェンダなどポジション）の学習が中心となっており、当初の期待に当たっていない。というのも、これはエージェント間の積極的な相互作用に基づく協調とは言えないからである。実際、フォーメーションや役割のみを獲得しても、味方の動きや敵の動きの予測をもとに行う関係プレーなどは実行できないと言われる。相互予測に基づく協調のメカニズムは、人間の協調行動に特徴的であり、これを機械学習の方法として枠組化することは、マルチエージェント学習にとって非常に重要である。そこで、人間の初心者にサッカーを指導する際の指導法に基づいた学習方法を、強化学習の枠組を用いてサッカーエージェントの学習に応用することを試みる。

第1章では、マルチエージェントシステムにおける学習の問題と、サッカーエージェント研究の現在の傾向について紹介する。そして、実際のサッカーとサッカーエージェントを比較したとき、サッカーエージェント研究で考慮されてこなかった戦術という要素がマルチエージェントの協調学習研究にとっても重要であることを述べる。

第2章では、人間の初心者がサッカーを学ぶ方法について述べる。サッカー指導者 Hughes の方法では、フルフィールドでの総合的プレーを練習させる以前に、プレーを領域に分けて領域の広さと性質に応じた技術を学習させる。そして、サッカーの戦術決定とは、他プレイヤーの行動の予測と

状態のリスク計算の最適化である，としている．総合すると，

1. フィールドのどの部分を使って練習するのかを設定する
2. 練習には何人のプレイヤーが参加するのかを設定する
3. 単純な練習内容の設定する
4. 行動の成功率に基づく判断をする
5. リスク計算に基づく判断をする

初心者にフルフィールドで4.や5.を実行させることは不可能であるから，1.,2.,3.のように，ゴールに至るまでのプレーを構成する小さなサブ課題を与えることで局面に応じた適切な協調プレーを学習させる．

本研究では，以上の学習方法に基づきエージェントプログラムを構築し，協調戦術を学習によって獲得する過程をシミュレーションする．エージェントを設計するにあたり，予測部には条件付確率を，状況に関するリスク計算の学習部分には，適応学習を導入する．具体的にはエージェントは，

1. 状況をグリッドにより離散状態変数化する，
2. 状態間の条件付確率をプレーの経験により学習する，
3. 与えられた課題の達成（もしくは失敗）に基づき，経験済みの状態の効用値を増減させ，最適にできるだけ近いリスク計算を学習する

状態間に定義される条件付確率は初期値としてランダムに与えられるが，プレーにおける状態遷移に基づき更新され，敵及び味方の振舞いについての情報を与える．各々の状態は初期においてランダムな効用値を持つが，プレーの過程でチームの課題達成に寄与した状態は報償を得る．シミュレーションにおける，ミニゲームの設定および，そこでエージェントに与えられる課題については以下に述べる．

第3章では， 3×4 のグリッドで攻守3対2の環境でシミュレーションについて述べた． 3×4 のグリッドによりエージェントは環境を離散状態変数化する．3対2という設定により，数的優位が局所領域で達成されている状態でプレイヤーが実行すべき協調戦術の学習が促せる．図1, 図2におけるグリッドの底辺の太線はエンドラインである．エージェントに与えられた課題は，

3人の攻撃側: エンドラインにボールを到達させること

2人の守備側: 攻撃を阻止しフィールドからボールをサイドアウトさせること

である．これが達成されたミニゲームの遷移状態の効用値を強化することで，リスクの学習を行う．

結果として，本エージェントは学習によって，図1, 図2のような小人数による基本戦術を獲得した．図1,2中の矢印は，プレイヤー，及び，ボールの移動を意味する．また，オープンスペースを使ってやり取りされるパスが学習によって増加し，それにつれて，攻撃側の勝率が有意に増加した．また，強化学習の代表的方法であるQ-learningとの比較の結果，マルチエージェント動的環境下の学習に対する本研究の頑健性が示された．

第4章では，第3章で提案した学習方法をより一般的な連続空間での学習環境に拡張し，マルチエージェント協調学習の方法としての一般性を示した．具体的には，エージェントは自分を中

心とした相対グリッドにより環境の情報を状態変数化した (図 3)。シミュレーション環境として RoboCup 標準サッカーサーバを用いた。

既存のサッカーエージェント YowAI の基本スキルをベースとし、意思決定部分を本研究の予測とリスク計算に基づく方法によるもので置き換えた。ミニゲームに参加するエージェント数は、シミュレーション 1 と同様に 3 対 2 とする。

計算量爆発を防ぐため、適当な情報の簡略化を行った。visibility という、ボールと敵の位置関係により定義される空間の安全指標と、味方と敵のどちらの数が優位か、の指標に基づき、各グリッドを 3 つの属性値 (有利, 不利, どちらでもない) に分類する。

状態の効用値の学習については、第 1 章と異なり、相対グリッドによって状態変数化されており、報酬を与えるための明確な課題達成の終端条件がない。逐次報酬として、味方間でパスが通ったときの状態系列に対して報酬を与える。

比較として、ベースとなるエージェントによる同様の環境でのシミュレーションを行った。ベースエージェント YowAI は、パスが出せる味方を選んでその方向にパスを出す、という行動をする。

シミュレーションの結果、攻撃側のボール支配時間について比較を行った。300 秒のミニゲームを 500 回繰り返した結果、学習エージェントのボール支配時間は、ベースエージェントの支配時間を有意に上回った (図 4 参照)。これは、状態表現に visibility というパスに関する空間の安全指標を導入したことにより、安全な空間がどこにできるかについて予測ができるようになっていたからだと考えられる。

以上より、本研究は、まず、グリッド上での局所領域でのミニゲーム環境において、空間的に俯瞰図として与えられる状態をもとにしたエージェントやボールなどオブジェクトの予測とリスク計算の学習を行い、協調戦術の獲得が可能であることを示した。ここで協調戦術とは、うまくパスを通すために都合の良い位置を探すというのではなく、常に、他エージェントの行動を予測し、評価しながら、自らもホームポジションなどに拘束されず移動することで、オープンスペースを介したパスなどを可能にするような、エージェント間の行動の相互予測に基づく行動であり、人間がサッカーを行う際の意味決定に近いものである。以上の方法は、動的環境下の学習として頑健であることが示された。

次に、より物理的な環境、つまり連続空間でシミュレーションするにあたり、相対グリッドを導入した。これにより、 3×4 の絶対グリッドでは、主にオブジェクトがどこにいるか、という情報を用いて意思決定を行っていたものが、フィールドの visibility の概念により、オブジェクトの予測から、空間の予測を行うようになった。この予測に基づく意思決定が、中盤でのボール支配時間を増やすことを示した。

以上の予測とリスク計算の学習に基づく意思決定により、従来の機械学習では実現できなかった、相互予測に基づく協調学習を可能にすることを示した。

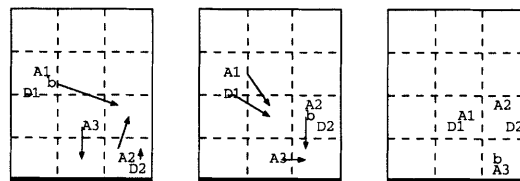


図 1: 壁パス

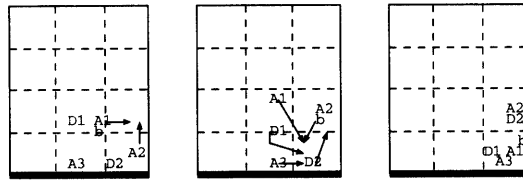


図 2: ワン・ツー・パス

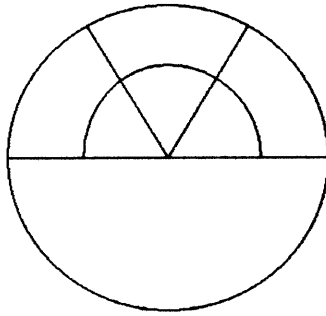


図 3: 相対グリッド化

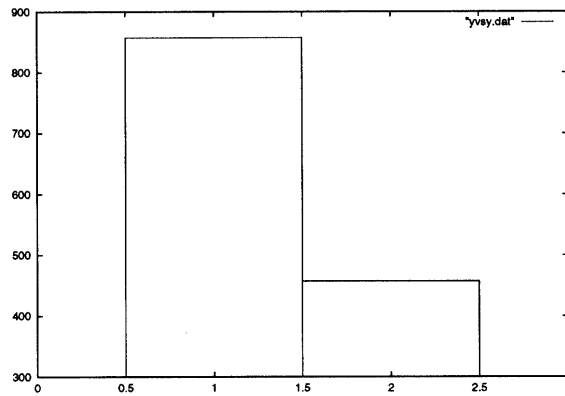


図 4: 左: 学習エージェントのボール支配時間, 右: 既定エージェントのボール支配時間