The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

実時間連続状態空間マルチエージェント意思決定に対する局面評価関数の設計について

On the design of state value functions
for real-time continuous-state space multi-agent decision making

中島 智晴 *1 Tomoharu Nakashima
五十嵐 治一 *2 Harukazu Igarashi
秋山 英久 *3 Hidehisa Akiyama

*1大阪府立大学 Osaka Prefecture University
*2芝浦工業大学 Shibaura Institute of Technology
*3福岡大学 Fukuoka University

This paper presents an overview of value function representations and construction for RoboCup soccer simulation. Since RoboCup has several characteristic features such as multi-agent system, noisy environments, and dynamic decision making, it offers a more realistic environment for the decision making in multi-agent system research.

1. はじめに

ゲーム AI の研究は，計算機の性能向上や深層学習の登場により急速な発展を見せている．その結果，囲碁や将棋をはじめとして，人間に勝利する AI プレーヤの登場が今や当たり前のようになってきた．ゲーム AI を構築するためには，センサ情報処理や意思決定など，社会的課題を解決するロボットや知的情報処理に転用できるサブ課題を解決し，システム統合する必要がある．このため，単純な課題であるトイプロブレムに比べてゲーム AI をベンチマークとする研究は実世界への応用可能性を広げるという意味で重要である

RoboCup サッカーシミュレーション 2D リーグは，約 20 年にわたる RoboCup の歴史の中で最も古いリーグの一つである．サッカーを題材とした仮想シミュレーションである一方で，複雑なマルチエージェント問題の研究プラットフォームとしても利用することができる．表 1 に，RoboCup サッカーシミュレーションと囲碁や将棋との違いをまとめる．これほどの制約がプレイヤに課されているゲームは他にはなく，RoboCup サッカーシミュレーションを対象としたゲーム AI の研究は，より実世界応用につながるものになること考えられる．

表 1: RoboCup と囲碁・将棋との違い

特徴	RoboCup	囲碁・将棋
情報	不完全	完全
思考時間	リアルタイム	ターン制
プレイヤ数	1 チーム 12 人	1 チーム 1 人
状態空間	連続	離散
行動空間	連続	離散

RoboCup サッカーシミュレーションのゲーム AI を構築するためには，センサ情報の処理，ボール処理の行動決定，次サイクルのセンサ情報を得るための行動，行動実施のためのミクロアクション決定，他エージェントを支援するための情報提供など，幅広い情報処理と意思決定機構が必要になり，数多くの人工知能的課題を解決しなければならない．このような複雑性もあり，RoboCup サッカーの知見を基にした在庫管理ロボットや深層強化学習モデルの発展につながっている．また，将来必要とされる集団ロボットの協調行動や自動運転における AI 部分の発展に大きく寄与すると考えられる．本論文では，その中でも意思決定部分に焦点を当て，RoboCup サッカーシミュレーションにおける意思決定メカニズムの研究を解説する．

連絡先: 中島智晴，大阪府立大学大学院，599-8531 大阪府堺市学園町 1 − 1，072-254-9351，tomoharu.nakashima@kis.osakafu-u.ac.jp

2. RoboCup サッカーシミュレーション

RoboCup は，「2050 年までにサッカーでロボットが人間のチャンピオンチームに勝利する」という目標を掲げて立ち上がった国際プロジェクトである．RoboCup サッカーには，いくつかのカテゴリーがある．本研究では，実機を使用せず，2 次元空間の仮想空間内で競技を行うサッカーシミュレーション 2 D リーグを対象とする．

2.1 構成

RoboCup サッカーシミュレーションは，仮想サッカーフィールド上で競技を行う，実機ロボットを使用しないリーグである．サッカーサーバ，サッカープレイヤ（以降プレイヤ），コーチ，モニタから構成される．この様子を図 1 に示す．サッカーサーバはサッカーフィールドの全ての情報を保持しており，プレイヤからのアクションコマンドを受信して次サイクルのフィールド状態を計算し，更新する．更新されたフィールドの情報を各プレイヤに送信する．プレイヤは，サーバから受け取られたフィールド情報に基づいて状況を判断し，次サイクルのアクションコマンドをサーバに送信する．プレイヤ間，コーチとプレイヤ間の通信はサッカーサーバを介してのみ許可されており，直接通信することは許されていない．1 ゲームは前後半それぞれ 3000 サイクルであり，1 サイクルの長さは 0.1 秒である（したがって，1 試合の長さは延長戦がなければ 6000 サイクル=5 分）．

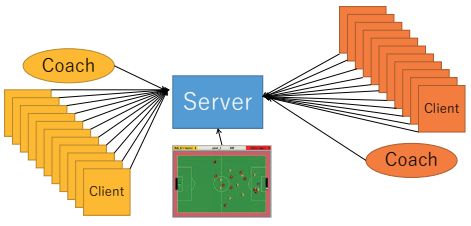

図 1: RoboCup サッカーシミュレーションの構成

2.2 プレイヤの行動決定

ここで，プレイヤの行動とアクションコマンドの違いについて述べる．特に，ボールを持っているプレーヤの意思決定に焦点を当てて説明する．プレイヤはフィールド状態に基づいて次サイクルの行動を決定するが，行動の意思決定には大きく分けて二つの層がある．まず，意思決定の第 1 層では，パス，ドリブル，シュート，ホールドなどのマクロ行動を決定する．これらの行動にはパラメータが必要である．

例えば，パスの場合は，パスを受け取るプレイヤ番号と受け取りが，ドリブルの場合には，目標位置がパラメータとなる．ドリブルでは，まずボールを蹴りだした後，次ボールを蹴る位置がパラメータとなり，シュートはゴールのどの位置に向ってボールを蹴るかがパラメータとなる．また，第 1 層で決定された行動に基づいて，第 2 層では，その行動とパラメータを達成するためのアクションコマンド（キック，ダッシュ，ターンなど 5 種類）とそのパラメータを計算する．これは，サーバ内であらかじめルールによって決められているボールやプレイヤの物理計算を逆算することで求められる．キックのパラメータは蹴る方向とその力，ダッシュは走る方向とその力，ターンはターンする量がパラメータとなる．プレイヤはアクションコマンドとそのパラメータをサッカーサーバに送信する．

本論文で取り上げる行動の意思決定は，この第 1 層における意思決定プロセスのことをいう．

3. 行動連鎖生成システム

行動連鎖システムは Akiyama ら [1] によって最初に導入され，他のチームも採用している．行動連鎖生成システムの目的は，ボールを保持しているプレーヤが次の行動を決定するための方針を決定することである．方針が定まれば，それに沿うような次の行動を決定することができる．

行動連鎖生成システムは，大きく 3 つに分けることができる．探索木の生成，ノードの評価，木探索アルゴリズムである．このうち，木探索アルゴリズムはゲーム AI で広く使用されているモンテカルロ木探索や系統的探索アルゴリズムが適用できる．RoboCup サッカー（もしくはサッカー AI）固有の要素が必要なのは探索木の生成とノードの評価である．ノードの評価はサッカーフィールドがどれほどスコアにつながっているかを表すもので，局面評価とも呼ばれる．

その評価関数中のパラメータを，人間の専門家同士の棋譜を用いて学習させる教師あり学習により，プロレベルまで棋力を向上させることができる．現在では，階層型のニューラルネットワークモデルを用いて評価関数を近似し，ディープラーニングの手法を取り入れることが主流になっている．特に，評価関数を作成するのが難しいとされた囲碁で大成功しており，チェスや将棋でも試みが始められている．

4. 局面評価モデルの構築

局面評価の基準は，チーム戦略によって変化する．例えば，シュートできることを目的とする戦略を考えている場合には，シュートという目標状態に近いが同課の観点から評価値を設定する．また，シュートまでの中間状態を目標とするチーム戦略を立てることも可能であり，例えば，ボールを敵ペナルティエリアに運ぶことを目的とするチーム戦略を考えている場合には，シュートできるかどうかは考慮せず，とにかくボールを敵ペナルティエリアに運ぶために有利な状態かどうかの観点から評価値を割り当てることになる．

4.1 特徴量の線形和による表現

RoboCup では，各プレーヤは視野範囲内にある他プレーヤの位置や速度，自己位置を知るためのフィールド外に設置されたランドマークの位置が，数値情報として利用可能である．ただし，自分から遠くにあるオブジェクトの情報にはノイズや欠損などの外乱が入り，不正確になる．ノイズや欠損，視野範囲外のブレーヤやボールの情報は他プレーヤからの声かけで補完されることもある．センサ情報にはノイズが大量に含まれており，プレーヤは正しくフィールド状態を把握することは非常に困難である．このような条件下でフィールド状態を評価する．

フィールド状態を評価する最も単純な方法は，フィールド状態を表現している特徴量をいくつか用意し，それぞれを重み付けながら足し合わせることである．この方法は，特徴量の線形和と呼ばれる．特徴量として，センサ情報から得られる敵プレーヤの位置やボール，自プレーヤの位置や速度などを用いたり,「近くに敵がいる」「パス可能な味方プレーヤが近くにいる」などのような，人間がサッカーの知識を導入して作成する特徴量も考えられたりする．後者の特徴量は特に，ヒューリスティクスと呼ばれることもある．

特徴量の線形和でフィールド状態を評価する場合，各特徴量に対応する重みをどのようにして決定するかという問題が残る．単純にはランダムに決定する方法や人間の知識を利用して決定する方法が考えられるが，限界や欠点も存在する．

4.2 ニューラルネットワークモデルによる表現

前節で述べた特徴量の線形和はパーセプトロンと見なすことができる．この見方を発展させて，特徴量を入力，評価値を出力とする階層型ニューラルネットを用いてフィールド状態の評価値を求めることもできる．

局面評価値計算モデルとしてパーセプトロンやニューラルネットワークを用いることで，機械学習の枠組みを用いた，経験データを用いたモデルの構築が可能となる．経験データの種類によって，評価値計算モデルを教師あり学習や強化学習により求めることができる．なお，教師無し学習により評価値計算モデルを構築する研究はあまり行われておらず，研究は進んでいない．次章では，教師あり学習と強化学習を用いたフィールド状態計算モデルの構築について述べる．

ニューラルネットワークの特徴量として，数値情報ではなく画像情報を用いる方法も考えられる．Pomas ら [2] は，フィールド状態を数値情報ではなく画像で入手し，その画像からチームにとって有利かどうかを判断する関数を構築している．

5. おわりに

本稿では，RoboCup サッカーシミュレーションにおける局面評価値計算モデルについて述べた．局面評価関数を構築する方法について整理し，今後の研究展望を考慮するうえで重要となる情報を与えることを目的とした．

参考文献

[1] H.Akiyama, S.Aramaki, T.Nakashima, ”Online Cooperative Behavior Planning Using a Tree Search Method in the RoboCup Soccer Simulation,” *Proc. of INCoS 2012*, pp.170-177, 2012.

[2] T.Pomas and T.Nakashima, ”Evaluation of Situations in RoboCup 2D Simulations using Soccer Field Images,” *Proc. of RoboCup Symposium*, 6 pages, 2018.