社団法人 人工知能学会
Japanese Society for
Artificial Intelligence

人工知能学会研究会資料
JSAI Technical Report
SIG-Challenge-B201-2 (5/4)

エージェントの意思決定における振動低減

Decreasing Oscillations in Decision Making of Agents

成本 洋介† 中島 智晴† 秋山英久‡

Yosuke NARIMOTO† Tomoharu NAKASHIMA† Hidehisa AKIYAMA‡

大阪府立大学† 福岡大学‡

Osaka Prefecture University† Fukuoka University‡

tomoharu.nakashima@kis.osakafu-u.ac.jp, akym@fukuoka-u.ac.jp

Abstract

In this paper, we propose a method that decreases the number of the oscillations in decision making by adjusting the importance of the previous decision making using a weighting function so that the agent tends to continue the previously selected action. With some numerical experiments, it is confirmed that our proposed method decreases the oscillations in decision making. We also investigate how the propesed method influences the performance of the overall team strategy.

1 はじめに

近年，動的で不確実な状況下におけるエージェントのプランニングに関する研究が幅広く行われている．例えば，山田は [1] において，ソフトウェアエージェントにおけるプランニングの研究事例としてインターネットエージェントなどを紹介し，プランニングの研究課題について，不確実な状況に対応することや，任意時間で実行可能なプランを出力することを挙げた．神尾ら [2] は，複数のロボットによる協調作業のための経路プランニングに対して，ランダムサンプリングを用いてサブタスクに必要なサブゴールを自動的に生成する経路プランニングアルゴリズムを提案した．また，エージェント同士の協調行動だけでなく，対戦型ゲームにおける人工知能エージェントの意思決定に関する研究も注目を集めている．例えば，Onieva ら [3] は，カーレーシングシミュレーションにおいて，後ろからくる対戦車両に対して，ファジィ制御を用いたブロッキング動作を行うカーエージェントを開発した．DeLooze ら [4] は，パックマンゲームにおいて，パックマンエージェントがどのようにゲームをプレイするかを学習するのに，ファジィQ 学習が有効であることを示した．

本論文では，RoboCup サッカー 2D シミュレーションにおいて，エージェントが前のサイクルで選択した行動を次のサイクルの意思決定で考慮することで，エージェントの意思決定の振動を低減する手法を提案する．本論文では，意思決定の振動を，前のサイクルで実行しようとしていた行動に対して，次のサイクルで違う行動を選択することと定義する．意思決定の振動が発生した場合，エージェントは時間を無駄に消費してしまうだけではなく，チーム全体を不利な状況にしてしまうという問題がある．提案手法では，本論文で使用されるエージェントの意思決定において，選択された行動の評価に対して修正式を加えることにより，意思決定の振動を低減する．

2 エージェントモデル

2D シミュレーションリーグにおいて，エージェントがどのように意思決定を行うべきかということは戦略開発における重要な要素である．例えば，Ma ら [5] は，専門知識のデータベースであるプランプールの概念を提案し，エージェントの協調行動のための方策探索計画法を適用した．Gspandl ら [6] は，自然言語のテキストから領域知識を自動的に抽出する手法を提案し，その手法をサッカーエージェントの意思決定に適用した．

本論文で使用するエージェントは最良優先探索を用いて，行動連鎖 [7] と呼ばれる木構造を構築することで意思決定を行う．以下に行動連鎖生成のアルゴリズムを示す．

Step 1: 現在の状態をルートノードに入力する．

Step 2: ノードにおいてエージェントが実行可能な，候補となる行動を生成する．
(pass, dribble, shoot など)

Step 3: 生成された行動の評価値 e を算出し，行動によって達成される状態とともに子ノードに追加する．

Step 4: ノード数が設定された最大数に到達していれば終了する．そうでなければ，e が最大となるノードを選択して，Step 2 へ戻る．

まず，エージェントが現在置かれている状態をルートノードに入力する．次に，ノードに入力された状態おいて，実行可能な行動の候補となる行動を生成する．生成された行動に対して評価値 e を算出し，行動によって達成される状態とともに子ノードに追加する．ノードが追加されるごとに，e が最大となるノードを選択し，ノードにおける状態から再び実行可能な候補となる行動を生成する．これを繰り返すことで，ノード数があらかじめ設定された最大数に到達するまで探索木を成長させる．このとき，葉ノードにおいて以下の条件のいずれかを満たす場合は，その葉ノードでの子ノードの生成は行わないものとする．

- 木の深さがあらかじめ設定した値を越えた．
- ノードに入力された状態から行動が生成できない．
- 行動連鎖の終了条件に設定されている行動 (shoot) が生成された．

構築された木構造のルートノードから葉ノードまでのノード列をつなげると，ある行動連鎖が得られる．最も評価値の高い行動連鎖を選択することで戦略的によいと考えられる行動連鎖を実行することが可能である．

行動連鎖の例を図 1 に示す．図 1 において，ボールを持った 10 番のエージェントは次のような行動連鎖を生成する．まず，10 番のエージェントが 7 番に pass を行う．pass を受け取った 7 番は dribble で前進をして，その後 9 番に pass を行う．pass を受け取った 9 番は相手ゴールに shoot を実行する．10 番のエージェントは以上の流れを考慮して，7 番への pass を実行する．

行動連鎖を生成することで，エージェントは数手先の状況を考慮してより戦略的価値の高い行動を選択することが可能である．しかし，サイクルごとにそれぞれの状況に対する行動連鎖の評価値を計算するため，サイクルが進むにつれて周囲の状況が変化し，最も評価値の高い行動連鎖が不適切に変化する場合がある．エージェントは通常，pass や dribble などの行動に 1, 2 サイクルの準備を必要とするため，選択された行動連鎖が変化すると，行動の準備がサイクルごとに切り替わってしまう．例えば，100 サイクル目で pass を選択し，pass の準備を始めたのに，101 サイクル目で dribble が選択されると，エージェントは pass の準備を止めて dribble の準備を始める．この様に，選択された行動が変化し，エージェントが無駄なサイクルを過ごしてしまうことを本論文では意思決定の振動と定義する．詳細な定義は次章で述べる．

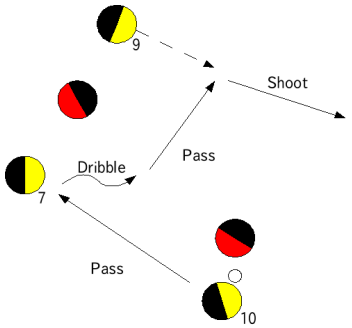


Figure 1: 行動連鎖の例

3 意思決定の振動

3.1 振動の定義

本論文では，エージェントの意思決定の振動を以下のように定義する．エージェントがボールを 2 サイクル以上連続して保持しているときに，連続した 2 サイクルで pass などの目標プレイヤが変更される場合．または，選択した行動の目標ボール座標が，前サイクルで選択した行動のそれと比べてしきい値となる距離以上ずれていた場合とする．ただし，行動のカテゴリが hold から，dribble や pass といった他のカテゴリの行動に変更した場合は，振動ではないものとする．hold とはボールを保持したままその場に止まる行動である．hold は dribble や pass などと違い，最終的には他の行動に変更しなければならないため，hold から他の行動に変更される場合は意思決定の振動から除外する．

振動と判定するための目標位置のずれについては，理論的なずれのしきい値を算出するのは困難である．そのため，ボールの目標座標の距離のずれのしきい値を決定するために以下のような予備実験を行う．

まず，エージェントを 10 回試合させる．ここで，エージェントがボールを 2 サイクル以上連続して保持しているとき，行動の目標座標が連続した 2 サイクルでどの程度の距離ずれるのかを計測する．ただし，目標となるプレイヤが変更された場合は距離のずれに関わらず意思決定が振動したとみなされ，除外されるものとする．表 1 に計測の結果を示す．ここで，距離は x 軸距離の 2 乗と y 軸距離の 2 乗の和とする．表から，0.6-0.7 の区間までは距離のずれの頻度が小さくなりつづけているが，0.7-0.8 の区間において再び頻度が大きくなっていることが分かる．これを考慮し，本論文では 0.75 を距離のずれをしきい値とする．

3.2 振動の事例

図 2 に意思決定の振動の事例を示す．図 2 の (a) において，左下の赤色の円で囲まれたエージェントは，右上の黄色の円で囲まれたエージェントへの pass を次の行動と決定している．そのために，首や体の向き，kick の準備をしなければならない．しかし，1 サイクル後の (b) では，周囲の状況が変化したため，右下の黒色の円で囲まれたエージェントに pass の対象を変更している．これに伴って，改め

Table 1: 目標位置の距離のずれ

距離 $(x^2 + y^2)$ の区間	ずれの頻度
0.0-0.1	518
0.1-0.2	103
0.2-0.3	54
0.3-0.4	25
0.4-0.5	23
0.5-0.6	22
0.6-0.7	5
0.7-0.8	13
0.8-0.9	13
0.9-1.0	34
1.0-1.1	15
1.1-1.2	10
1.2-1.3	10

て首や体の向きの変更，さらにキックの準備をしなければならない．さらに，その次のサイクルである (c) では，パスを実行不可能と判断し，自分で dribble をしようとしている．従って，再び首や体の向き，キックの準備を始めなければならない．このように，状況が少し変化するだけでエージェントの意思決定は大きく変更される．結果として，エージェントは各行動に対する準備のためサイクルを無駄にしてしまい，敵に近づかれることにより，ボールを奪われやすくなってしまうという問題が発生している．

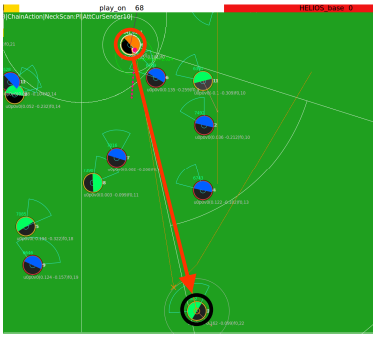
3.3 振動の統計的調査

意思決定の振動が試合中にどの程度発生するのかを調査するため，エージェントに 20 回試合をさせ，以下の項目に関して調査を行う．

- 意思決定を行った回数全体
- 2 サイクル以上連続して意思決定を行った回数
- 意思決定が振動した回数
- 意思決定全体に対する振動の割合
- 連続した意思決定に対する振動の割合

ここで，2 サイクル以上連続して意思決定を行った回数を計測するのは，意思決定の振動は連続サイクル上で定義されているためである．結果を表 2 に示す．表 2 の結果から，連続した意思決定に対して，高い頻度で振動が発生していることが分かる．また，これ以降は，意思決定の振動の割合は，連続した意思決定における意思決定の振動回数の割合とし，意思決定を行った回数全体は考慮しないものとする．

(a) 67 サイクル目



(b) 68 サイクル目

(c) 69 サイクル目

Figure 2: 振動の事例

Table 2: 振動の発生回数

意思決定の回数	18119
連続した意思決定の回数	1926
意思決定が振動した回数	1389
全体に対する割合	0.07666
連続したものに対する割合	0.7212

4 提案手法

4.1 提案手法のアルゴリズム

本論文で対象とするエージェントは，行動連鎖を生成することで意思決定を行う．行動連鎖の生成過程では，候補となる行動に対して評価値を算出し，最も評価値の高い行動連鎖を採用している．そこで，提案手法では，前サイク

ルでの行動連鎖を考慮して評価値に修正を加えることで，意思決定の振動を低減する．以下に，提案手法のアルゴリズムを示す．

Step 1: 現在の状態をルートノードに入力する．

Step 2: ノードにおいてエージェントが実行可能な，候補となる行動を生成する．
(pass, dribble, shoot など)

Step 3: 生成された行動の評価値 e を算出する．

Step 4: 前の行動連鎖が記憶されていれば，e を修正する．そうでなければ，e は修正しない．

Step 5: 生成された行動によって達成される状態と e を子ノードに追加する．

Step 6: ノード数が設定された最大数に到達していれば e が最大となる行動連鎖を決定し，Step 7 へ移動する．そうでなければ，e が最大となるノードを選択して，Step 2 へ戻る．

Step 7: 前の行動連鎖が記憶されていないか，前の行動連鎖での最終目標地点との距離にしきい値以上のずれがある場合，生成された行動連鎖を前の行動連鎖に上書きして記憶する．

Step 8: 前の行動連鎖が生成されて一定サイクル経過している場合，削除する．

提案手法では，エージェントが候補となる行動を生成するとき，生成された行動の評価値 e を修正することによって，意思決定の振動を低減する．前の行動連鎖が記憶されている場合，後述する修正式からそれぞれの行動の評価値 e を修正する．生成された行動に対して，前の行動連鎖における同一の深さの行動を使用して，それぞれ修正式を適用する．同一の深さのものがない場合，最後の行動を使用するものとする．記憶された行動連鎖がない場合，評価値の修正は行わない．

最終的な評価値を計算した後，最も高い評価値を持つ行動連鎖が意思決定として選択される．このとき，前の行動連鎖が記憶されていなかった場合，新たに生成された行動連鎖を記憶する．また，新たな行動連鎖の最も深い行動の目標位置が前サイクルの行動連鎖のものとしきい値以上の距離のずれがある場合，新しく生成された行動連鎖を採用し，実行する．これは，しきい値以上の距離のずれがあるのに高い評価値を示している行動連鎖は，状況の変化に対応したよりよい行動連鎖であると考えられるためである．記憶された行動連鎖は，行動連鎖が生成されてから終了するまでにかかると予測されたサイクル数のあいだ持続するものとし，そのサイクルを過ぎた場合，基準となる行動連鎖は削除される．

4.2 評価値の修正

評価値 e は以下の式によって修正される．

$$e' = e \times exp(-k\frac{\| \boldsymbol{p} - \boldsymbol{q} \|^2}{(1+(n-m))^2}) \quad (1)$$

ここで，e' は修正後の評価値である．また，n と m はそれぞれ，現在の試合時刻と前の行動連鎖を選択したときの試合時刻である．$\boldsymbol{p}$は候補行動連鎖における行動のボールの目標位置，$\boldsymbol{q}$は現在実行中の行動連鎖における，同じ深さの行動のボールの目標位置である．さらに，k は時間と距離の影響を変化させるための非負のパラメータである．

この式は前の行動での目標座標を平均，時間の経過を分散としたガウス分布と類似している．これは目標座標との距離のずれが大きくなるほど評価値が小さくなり，行動が選択され難くなることを意味する．また，現在の時刻と前の行動の時刻が離れるほど，ガウス分布における分散が大きくなり，分布の山がなだらかになる．このことにより，試合時刻が経過するほど修正式の影響は小さくなり，ボールの目標位置が前の行動と大きく異なった行動も選択されるようになる．なぜなら，試合時刻が経過することで状況の変化が大きくなり，距離のずれが大きくても，より高い評価を持つ行動連鎖が採用される必要が出てくるからである．

5 数値実験

5.1 実験設定

使用するエージェントは，RoboCup2011 の決勝戦で使用された HELIOS2011 に提案手法を組み込んだものとする．行動連鎖の最大の深さを 3 とし，生成される行動数を 500 とする．数値実験では，意思決定の振動が低減されるかどうかを調査するため，提案手法のアルゴリズムを組み込んだチームと，提案手法のアルゴリズムを組み込んでいないチームを，それぞれ agent2d と対戦させる．agent2d[9] は，フリーで公開されているサンプルのチームである．HELIOS2011 と同じように意思決定に行動連鎖を用いており，いくつかの世界大会登録チームによって開発のベースとして利用されている．数値実験では，また，修正式のパラメータ k の値を変化させることで，時間と距離に対する重みが意思決定の低減にどのように影響するかも調査する．k の値を 0.1, 0.5, 1, 1.5, 3, 5, 10, 50, 100 と変化させて，20 回ずつ試合を行い，エージェントの意思決定の振動の回数を計測する．

また，同様の実験を agent2d でも行う．行動連鎖の深さは 1 とし，生成される行動数を 500 とする．提案手法を組み込んだ agent2d とそうでない agent2d を対戦させ，意思決定の振動が低減されるかどうかを確認する．修正式のパラメータ k の値は 0.5, 1, 1.5, 3.0, 5, 10 と変化させ，10 回ずつ試合を行い，エージェントの意思決定の振動の回数を計測する．

5.2 実験結果

HELIOS2011 の結果を表 3，agent2d の結果を表 4 に示す．表 3 と表 4 から，両方のチームにおいて，提案手法を組み込んだ場合は，提案手法を組み込まなかった場合に比べて意思決定の振動の割合が大幅に減少していることが分かった．また，表 3 では，パラメータ k の値が 1.0 から 5.0 の範囲にあるとき，振動の割合が最も低くなっていた．ここで，k の値は大きくなるほど距離に対する重みが大きくなるので，距離と時間に対する重みが同じか，距離に対する重みがやや大きい状態が最もよい結果を出すことが分かった．表 4 では，k の値が大きくなるほど振動の割合が減少していた．しかし，k の値が大きくなるにつれて減少の度合いが減っているため，ある程度以上は大きくしても意味がないと考えられる．

5.3 性能評価

次に，意思決定の振動を低減した場合におけるチームの性能評価を行う．使用するチームは HELIOS2011 とし，その他の条件は数値実験と同様である．性能評価は人間のサッカーにおけるデータ分析 [8] でも取り入れられている，ボール保持率などに注目して行う．対戦させるチームは agent2d，AUA2D，Edinferno，HfutEngine，Photon，RMAS，WrightEagle の 7 チームとする．チームに提案手法を組み込んだ場合と組み込まなかった場合で，それぞれ 20 回試合をさせて性能を評価する．k の値は 1.0 とする．結果を表 5 から 9 に示す． ここで，太字になっている項目は，提案手法がある場合とない場合で，優位水準が両側 5%の t 検定で，平均に有意差があるとみなされたものである．また，表 5 において，振動の割合が 1.0 を超えるものがあるが，これは連続した意思決定に対して 2 回以上振動が発生する場合があるためである．表 5 では，提案手法を組み込むことによって，全てのチームに対して，意思決定の振動回数が減っている．このことから，提案手法は相手チームによらず効果があることが分かった．また，表 6 では，いくつかのチームで平均得点が低下していたが，表 7 では，いくつかのチームに対してボール支配率が向上していた．これは意思決定の振動が低減することでボールを保持しつづける行動が選択されやすくなり，ボール支配率は上昇しても得点に結びつかなくなってしまったためであると考えられる．また，表 8 と 9 の結果から，パスの成功回数よりもドリブルの成功回数が向上しており，提案手法はパスよりもドリブルに対して強く影響していることが分かった．

Table 3: HELIOS2011 の意思決定の振動の回数

k	意思決定回数	振動	割合
手法なし	1926	1389	0.7212
0.1	2677	791	0.2955
0.5	3130	709	0.2265
1	3634	594	0.1635
1.5	3661	741	0.2024
3	4047	619	0.1530
5	4085	643	0.1574
10	4414	966	0.2188
50	5264	1012	0.1922
100	4676	963	0.2059

Table 4: agent2d の意思決定の振動の回数

k	意思決定回数	振動	割合
手法なし	490	347	0.7082
0.5	1411	522	0.3700
1	1631	539	0.3305
1.5	1900	529	0.2784
3	2021	484	0.2395
5	2055	483	0.2350
10	2085	517	0.2480

Table 5: 意思決定の振動の割合

	提案手法なし	提案手法あり
agent2d	0.7212	**0.1635**
AUA	1.064	**0.4340**
Edin	1.051	**0.4221**
Hfut	1.255	**0.4983**
Photon	0.8118	**0.3147**
RMAS	1.082	**0.4284**
Wright	1.257	**0.4721**

Table 6: 得失点

	提案手法なし		提案手法あり	
	得点	失点	得点	失点
agent2d	3.1	0.6	3.3	0.5
AUA	**15.2**	0	11.6	0
Edin	**19.85**	0	17.15	0
Hfut	**19.25**	0.05	15.95	0.15
Photon	18.65	0	18.05	0
RMAS	**19.7**	0	16.7	0
Wright	4.8	0.8	4	1

Table 7: ボール支配率

	提案手法なし	提案手法あり
agent2d	0.6578	**0.6965**
AUA	0.4813	0.4996
Edin	0.6429	**0.6840**
Hfut	0.5909	0.6014
Photon	0.5454	**0.6037**
RMAS	0.7720	0.7761
Wright	0.4381	0.4272

Table 8: パスの成功回数

	提案手法なし	提案手法あり
agent2d	102.8	99.95
AUA	113.75	**127.75**
Edin	119.05	121.05
Hfut	139.35	141.65
Photon	113.35	105.6
RMAS	120.75	122.5
Wright	112.25	104.95

Table 9: ドリブルの成功回数

	提案手法なし	提案手法あり
agent2d	434.05	**482.8**
AUA	145.5	**172.4**
Edin	166.05	**221.9**
Hfut	110.85	**165.5**
Photon	170.7	**314.6**
RMAS	165.9	**232.85**
Wright	93.55	**140.6**

6 おわりに

本論文では，RoboCup サッカーシミュレーション 2D リーグにおいて，プレイヤエージェントの意思決定の振動を低減させる手法を提案した．提案手法によって，エージェントが意思決定を行うとき，前に選択された行動を考慮して意思決定の評価値を修正することで，前に選択された行動が選択されやすくなり，意思決定の振動の発生が低減された．また，数値実験によって，評価値を修正することで意思決定の振動が低減されていることを確認した．さらに，性能評価により提案手法が適用されたことによるチーム全体への影響を調査し，いくつかのチームに対して平均得点数は減ったものの，ボール支配率が上昇したことを確認した．本論文ではエージェントがボールを持つときの意思決定のみに着目していたが，ボールを持たないエージェントのインターセプト動作などに振動低減を拡張することが，今後の課題である．

参考文献

[1] 山田誠二，“ソフトウェアエージェントにおけるプランニング”，人工知能学会誌，page 623-628，2001.

[2] 神尾正太郎，伊庭斉志，“マルチエージェント協調作業のためのランダムサンプリングを用いた経路プランニングアルゴリズム”，電子情報通信学会論文誌，D, 情報・システム，pp. 250-260，2006.

[3] E. Onieva, L. Cardamone, D. Loiacono and P.L. Lanzi, “Overtaking Opponents with Blocking Strategies Using Fuzzy Logic,” In *IEEE Conference on Computational Intelligence and Games*, pp. 123-130, 2010.

[4] L.L. Delooze and W.R. Viner, “Fuzzy Q-Learning in a Nondeterministic Environment:Developing an Intelligent Ms.Pac-Man Agent,” In *IEEE Conference on Computational Intelligence and Games*, pp. 162-169, 2009.

[5] J. Ma and S. Cameron, “Combining Policy Search with Planning in Multi-agent Cooperation,” In *RoboCup 2008: Robot Soccer World Cup XII*, pp. 532–543, 2008.

[6] S. Gspandl, A. Hechenblaickner, M. Reip, G. Steinbauer, M. Wolfram and C. Zehentner, “The Ontology Lifecycle in RoboCup:Population From Text and Execution,” In *RoboCup Symposium 2011*, pp. 313-324, 2011.

[7] 秋山英久，“アクション連鎖探索によるオンライン戦術プランニング”，第 33 回人工知能学会 AI チャレンジ研究会，pp.23-28，2011.

[8] 森本美行，本田にパスの 36%を集中せよ ザック JAPANvs. 岡田ジャパンのデータ解析，文春新書，2011.

[9] agent2d http://rctools.sourceforge.jp/pukiwiki/index.php?agent2d