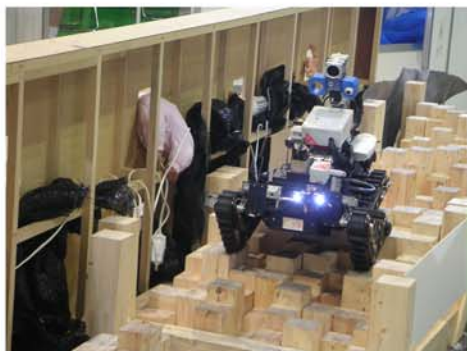


人工知能学会 第23回SIG-Challenge研究会



2006

© 2005 The RoboCup Federation.



2006年5月4日
西日本総合展示場（北九州市）
ジャパンオープン2006

目次

1. 分散処理環境での都市型水害シミュレータの検討, 鈴木 崇泰, 高橋 友一 (名城大学大学院)	1
2. 方策勾配法を用いたサッカーエージェントの学習 ~ フリーキックにおけるキッカーとレシーバ~, 中村 浩二 [1], 五十嵐 治一 [2], 石原 聖司 [3] (1 芝浦工業大学大学院, 2 芝浦工業大学, 3 近畿大学工学部電子情報工学科)	7
3. CITIZEN Eco-Be! League: bringing new flexibility for research and education to RoboCup, Rodrigo da S Guerra[1], Joschka Boedecker[1], Norbert Mayer[1,2], Shinzo Yanagimachi[3], Yasuji Hirosawa[3], Kazuhiko Yoshikawa[3], Masaaki Namekawa[3], and Minoru Asada[1,2] (1 Graduate School of Engineering, Osaka University, 2 HANDAI Frontier Research Center, 3 CITIZEN Co.)	13
4. 位置に基づく行動規則を実装した入門用サッカーエージェント作成キット OZED, 西野 順二 [1], 久保 長徳 [2], 下羅 弘樹 [3], 中島 智晴 [4] (1 電気通信大学, 2 仁愛大学, 3 福井大学, 4 大阪府立大学)	19
5. ニューラルネットワークによるゲームログからの行動獲得, 並川 直樹, 小野 玄基, 横山 智史, 高谷 将裕, 中島 智晴, 石淵 久生 (大阪府立大学)	25
6. ad-hoc ネットワークにおける音波を用いた端末間距離測定に関する一考察, 植村 渉 (龍谷大学)	29
7. マルチ全方位ビジョンシステムによる物体認識および自己位置同定手法, 井戸 大介, 前田 陽一郎 (福井大学大学院)	34
8. Inferring other's intention based on estimated state value of self, Yasutake Takahashi[1,2], Teruyasu Kawamata[1], Tom Tamura[1], and Minoru Asada[1,2] (1 Graduate School of Engineering, Osaka University, 2 HANDAI Frontier Research Center)	40

分散処理環境での都市型水害シミュレータの検討

Urban Flood Simulation on Distributed Computation Environment

鈴木崇泰 高橋友一

Takayashu Suzuki Tomoichi Takahashi

名城大学大学院 理工学研究科

Meijo University, Nagoya

ttaka@ccmfs.meijo-u.ac.jp

Abstract

Various reasons cause floods in cities. Not only bank damages but also localized burst may cause disasters in urban, especially people in underground city. Proper actions such as closing floodwalls can prevent water to flow into undergrounds. It is hard to simulate both flood flow and actions to them.

In this paper, we show a flood simulation based on diffusion equation model and discussion using experiments data using heavy rainfall disaster around Nagoya City.

1 はじめに

スマトラ地震による津波、河川の氾濫による洪水などに加え、近年は都市化により土地の雨水貯留、浸透量が低下し、短期間に多量の雨が河川などに流出しやすくなっている。その結果、都市機能に大きな被害をもたらす都市型水害が発生している[1]。事前に被害を抑える対策を検討するため、全国のいたるところで破堤や集中豪雨による洪水のシミュレーションができる事が望まれる。

地方整備局による河川での氾濫シミュレーションの表示[10]や、三重県尾鷲市を対象とした津波災害のシミュレーション結果などが公開されている[2]。都市域および地下空間を複数の貯留槽が連結して構成される空間とした貯留槽モデルを用い、福岡水害や京都における鴨川の溢水による浸水解析の報告がある[3]。国土交通省の「水情報国土」により個人レベルで雨量や川の水位情報を入手できる[4]。国土地理院の数値地図データを利用し、日本全国の地域で、50メートル四方単位の精度で堤防の決壊や河川の氾濫による洪水をシミュレーションするシステムが発表されている[5]。

都市型洪水では、排水しきれなかった水が地下街や地下室などへ浸水により、ドアが開かなくなる、停電するなど、被災者は気付いた時には避難できなくなる事がある。そのような都市型水害シミュレータに対する要求項目を以下に挙げる。

1. 任意の出水個所の指定 (破堤, マンホールからの出水, 水道管の破裂).
2. 救助活動 (修復工事, 排水ポンプの設置) の反映.
3. 避難先として建物.

平成 12 年 9 月に発生した東海豪雨のように、広範囲に被害が及ぶことがある (図 1)。高橋らは、洪水や地震などの災害時における避難勧告などによる人的損害への影響をふくめたシミュレータの構築を検討している[6]。

本論文では、分散処理環境下で構築した洪水シミュレーション結果の評価と分散処理の計算時間短縮の効果について報告する。



Figure 1: 東海豪雨による浸水の様子

2 洪水シミュレータの概要

本シミュレータは震災総合シミュレーションシステム (IDSS) のサブシミュレータとして開発した[7]。IDSS のサブシミュレータとして開発することで以下の機能を実現できる。

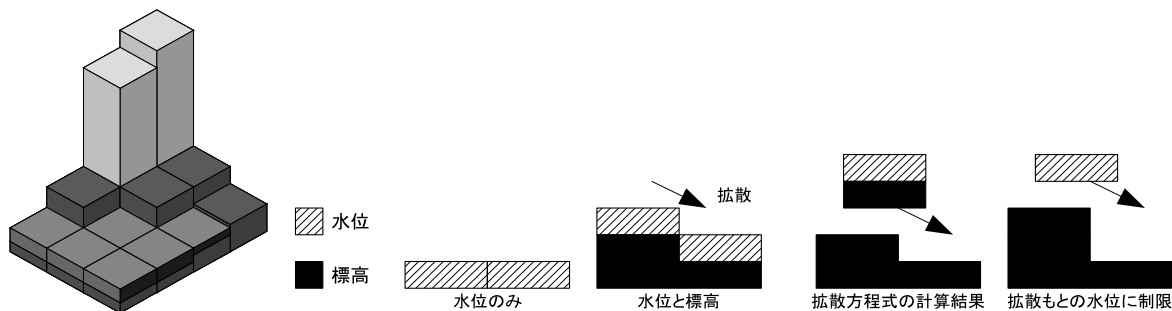


Figure 2: 格子オブジェクトによる標高表現と標高を考慮した拡散計算

1. 複数パソコンによる分散処理.
2. 共有時空間データ (SSTD) によるサブシミュレータ間の情報共有による
 - 火災など他の災害シミュレーションの計算結果や,
 - 市民や救急隊のエージェントの動きなどとの統合.

洪水シミュレーションには CFD(Computational Fluid Dynamics) などが用いられる. それらの方法では, 建物などの境界条件を規定する必要がある, 防水壁や排水ポンプの設置などの防災・救助活動を反映することが困難であるなどの問題がある.

2.1 格子オブジェクトによる計算

建物境界や防災活動を反映されたシミュレーションを実施するために, シミュレーション対象地域を格子状に区切り, 各格子を標高と水位のデータをプロパティとしてもつオブジェクトとして扱う. 水の移動は以下の拡散方程式の差分方程式に従うものとする.

$$\frac{\partial w_{i,j}}{\partial t} = \alpha \left(\frac{\partial^2 w_{i,j}}{\partial x^2} + \frac{\partial^2 w_{i,j}}{\partial y^2} \right) + S_{i,j} \quad (1)$$

ここで, α は拡散係数である. 正の $S_{i,j}$ は増水箇所を, 負は排水箇所を示し, その他の場所は 0 の値を取る.

$w_{i,j}$ は格子オブジェクト $P_{i,j}$ における土地の標高・建物と水位の和を示す (図 2). $w_{i,j}$ の変化においては, 水位を超えた拡散が起こらないように計算する. 格子サイズを建物の大きさ (10m 強) にする事により, ある $w_{i+1,j}$ が浸水していても, 隣接の $w_{i,j}$ に建物があるケースは, アドホックな処理で防水壁により浸水していない, 又は床上浸水であるといったシミュレーションが可能になる.

2.2 GIS データの用意

都市型の洪水シミュレーションでは, 一般の洪水シミュレーションに必要な標高, 河川のデータに加えて, 排水箇所の場所と量に関するデータと建物のデータが必要になる.

前者については, 国土地理院発行の数値地図 25000 にあるメッシュ標高ファイルにあるデータを補間し作成し

た.(図 3) 建物情報については, 住宅密度が統計データと同程度になるように自動生成されたデータを使用した[8].

```

- <!-- メッシュ標高 -->
- <MeshuHyoko id="MH23111000001">
  <Hyoko>0</Hyoko> //標高値
- <point id="PM23111000001">
  <CRS idref="JGD2000" />
  <position>492645.0000 126123.0000</position>//
  経度, 緯度
</point>
</MeshuHyoko>
.....

```

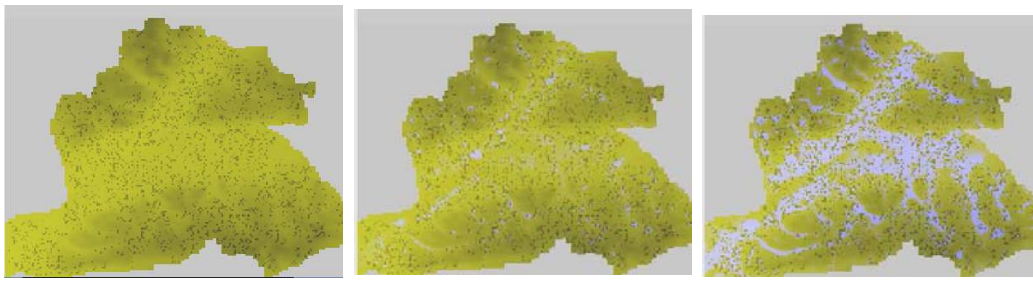


Figure 3: 数値データから作成した天白区の標高

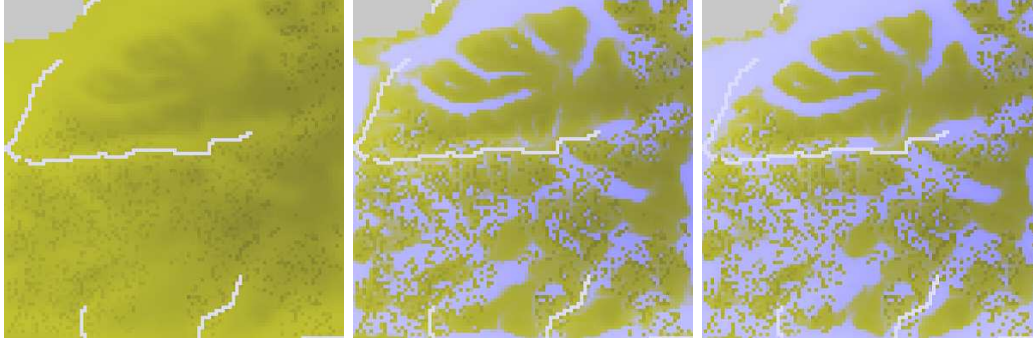
3 東海豪雨データとの比較

東海豪雨のデータを基づき, 名古屋市の天白区と野並地区を対象に集中豪雨のシミュレーション結果を評価した. 以下の条件でシミュレーションを行い, 図.4 に時間変化を示す.

1. インターネット上で公開されている豪雨時の 1 時間降水量や日降水量を利用する[9]. 公開されている降水量 (例えば, 時間最大雨量は 93mm) を, シミュレーションステップと格子面積で換算して値を用いた.
2. 排水ポンプの排水量は公開数値 ($18m^3/s$) を用い, 格子面積とシミュレーションステップ幅を考慮した値を用いる[11].
3. 一日をシミュレーション対象とし, 100 秒を 1 ステップとして 864 ステップの実行した. 格子サイズは, 1 辺の最大格子数が 100 になるように設定した.



(1) 天白区での集中豪雨の時間変化



(2) 野並地区での集中豪雨の時間変化

Figure 4: 集中豪雨シミュレーションの時間変化 (左から step=0,432,864. 薄い色が水を示す.)

3.1 評価と考察

東海豪雨における被害のデータを用いシミュレーション結果の評価を行う[12].

3.1.1 天白区

天白区 (約 $5.6\text{km} \times 7.4\text{km}$) を 75×100 の格子に分割した.

実験 1: 排水ポンプ (排水量 $S=-0.326$) を (50,25) と (50,50) の 2 箇所, それ以外の地点では, 降雨による増水量 ($S=0.00258$) を設定した. 1 台の PC (Pentium 4 (3GHz), 2GB メモリ) によるシミュレーション実行時間は 674[sec] であった.

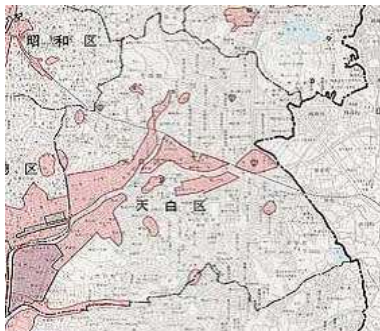


Figure 5: 東海豪雨における名古屋市天白区の被害状況

全格子に対する水位の割合の 6 時間毎の経過を表 1 に示す. 図 5 に示す天白区における実際の被害を,

12×14 の格子状に分割して浸水被害の割合をみると約 22% になる. 水位が 0.5 m 以上の格子で浸水家屋が発生したとすると, 最大時間降水量で 24 時間降り続いたシミュレーション結果は, 当然ながら実際の浸水被害に比べ広がった結果を示している.

実験 2: 2000 年 9 月 11 日における時間降水量を表 2 に示す[13]. このデータをもとに, 一日の平均降水量と一時間毎に降水量を設定した場合 (総降水量は同じ) のシミュレーションを行った. 天白区における浸水被害の割合は, 一日の平均降水量を用いた場合で 21.2%, 一時間毎の時間降水量を用いた場合は 24.4% となり実験 1 と比べ実際の浸水被害状況に近くなった.

Table 2: 2000 年 9 月 11 日 時間毎の降水量 (mm)

時刻	0	1	2	3	4	5	6	7
降水量	0	5	11	5	8	23	3	3
時刻	8	9	10	11	12	13	14	15
降水量	2	0	3	5	7	9	10	16
時刻	16	17	18	19	20	21	22	23
降水量	22	93	48	74	25	18	45	37

実験 3: 限られた機材を使用して, 効率のよい防災活動をする事が望まれる. 排水ポンプを標高の低い位置

Table 1: 時間あたり最大降水量 ($S = 0.00258$) 時の水位変化

水位	経緯時間					
	4	8	12	16	20	24
0.1m 未満	23.6 %	22.0 %	21.0 %	20.4 %	20.1 %	19.8 %
0.1m 以上 0.5m 未満	55.5 %	19.5 %	17.0 %	15.6 %	14.7 %	13.9 %
0.5m 以上 1m 未満	15.7 %	34.4 %	6.5 %	5.3 %	4.9 %	4.6 %
1m 以上	5.2 %	24.1 %	55.5 %	58.6 %	60.3 %	61.7 %

に変更してシミュレーションを行った結果を表 3 に示す。自明な事だが、水の集る標高の低い箇所に排水ポンプを設置した方が実験 2 より浸水被害が 0.2 % から 0.4 % 小さくなっている。

Table 3: 排水ポンプ設置位置と天白区浸水領域

排水ポンプ位置	一日平均降水量	時刻毎に変化
実験 1 と同じ	21.2 %	24.4 %
標高に低い場所	21.0 %	24.1 %

以上より、格子オブジェクトによる洪水シミュレーション結果は、実際に報告されているデータに近い事と、排水ポンプの設置場所の違いによる防災活動の結果を反映できることを示した。

3.1.2 野並地区

天白区に隣接し同じく浸水被害にあった野並地区 (約 $4.0\text{km} \times 4.0\text{km}$) を 100×99 の格子に分割し、排水ポンプ ($S=0.011$) を (50,25) と (50,75) の 2 箇所、それ以外の地点は降雨による増水 ($S=0.00258$) に設定した。シミュレーション実行時間は 829[sec] である。天白区と同様の降水条件で野並地区でシミュレーションを行い、浸水した箇所の割合を表 4 に示す。

東海豪雨における野並地区の被害状況を図 6 に示す[14]。一日降水量、一日の時間降水量の変化を用いたものが実際の被害に近い事がわかる。図 7 の上段に表 2 の時間降水量の時間変化を、下段に一日平均降水量と一日の時間降水量に対する格子点 ((x, y)=(14, 60)) 上における水位の変化を示す。図から、降水量に応じて水位が変化していることがわかる。

Table 4: 野並地区における水位 0.5m 以上浸水領域の割合

最大降水量	一日平均降水量	時刻毎に変化
50.1 %	23.6 %	21.0 %



Figure 6: 東海豪雨における野並地区の被害状況

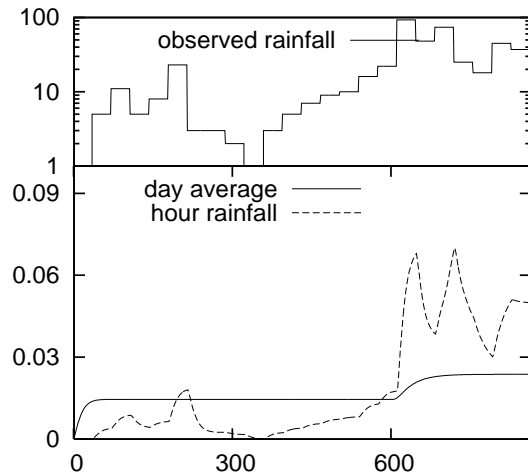


Figure 7: 東海豪雨時における時間当りの降水量とシミュレーション水位の時間変化 (一日平均降水量, 時間毎の降水量)

4 分散処理による高速化

図7に示すように、17時間後における降水量の増加に伴い急激に水位があがっている。都市型の水害においては、このような豪雨の観測と予測に基づく地下室からの迅速な避難が望まれる。そのためには、計算時間がかからない事が必要となる。一方で、2節で述べたように、建物を考慮したシミュレーションをするには、格子幅を建物の大きさまで小さくする必要がある。格子幅を小さくすると格子数を増加し、計算時間が必要になる。計算時間の増加を解決するために、IDSS上に地域分散による分散処理を行った。

Table 5: 各地図の分割数と実行時間(秒)

(1) 全体のサイズを一定

分割数	格子サイズ	1	2	4
港区	100 × 74	452	363	327
(港, 中川) 区	94 × 100	573	424	409
(港, 中川, 中村) 区	78 × 100	494	355	348
(港, 中川, 中村, 西) 区	61 × 100	417	342	321

(2) 地域分割されたサイズを一定

分割数(全体の格子サイズ)	1(100)	4(200)
港区	452	1,159
(港, 中川) 区	573	1,558
(港, 中川, 中村) 区	494	1,334
(港, 中川, 中村, 西) 区	417	1,253

スイッチハブを介してギガイーサーで同一仕様のPC(Pentium4 3GHz, メモリ 2GB)を使用して分散シミュレーション(分割数=1,2,4)をした時の結果を表5に示す。ステップ数は3節と同じく24時間で864ステップ数で、長い方の辺を図8に示すように方式で地域分散を行った。隣接領域間で情報の交換をするため糊代処理が必要になるオーバーヘッドを除くと、分割数(n)に応じて表5における値は

(1) 全体のサイズを一定のまま分割数を増やした時は、実行時間は $\frac{1}{n}$ に短縮される、

(2) 分割された地域のサイズが一定(全体のサイズは分割数に比例する)の時は、実行時間は n によらない

事が理想であるから、分散処理の効果はみえるもののオーバーヘッド部分に対し検討すべき点がある結果である。

5 議論

1時間毎の降水量や1日降水量を用い東海豪雨のシミュレーションを行った。1章に述べた要求事項のうち、任意の出水個所の指定、防災活動としての排水ポンプの設置を

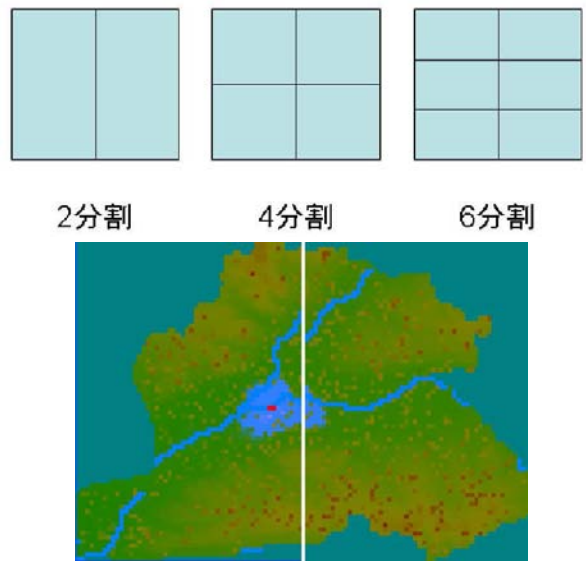


Figure 8: 地域分割と2分割時のりしろ処理

した結果を示した。それらを踏まえ、都市水害シミュレーションにおいて、

- 局所的な豪雨と発生する水位の上昇の関係：
格子オブジェクトと差分方程式による方法で、区単位の広域で実際の被害状況結果にあった値と、時間単位の降水量の設定により急激な水位の上昇がある事を示した。
- 建物などの影響を評価するための分解能と計算時間：
港, 中川, 中村, 西区がまとまった広さは、約13km × 21kmである。この領域を、格子サイズとして建物や地下街への入口として10mの分解能でシミュレーションするには、今回よりも一桁オーダーが大きい1300 × 2100の格子サイズが必要になる。
雨量計などセンサ情報から避難までの時間を10分、実験で使用した計算資源を前提にすると、リアルタイムで計算するのではなく事前に計算した結果を予報に使用する方式が現実である。

今後は、実際の水流の速さに合せたステップ幅と糊代幅の設定などシミュレーションの精度の向上に加え、戸外と建物内の間として境界を設定することによる建物内への水の侵入シミュレーションや土嚢などや排水など防災エージェント活動や建物内の避難行動などによる被害の軽減評価のシミュレーションの活用を検討する。

参考文献

- [1] http://www.sonpo.or.jp/disaster/flood/flood_01_4.html
- [2] <http://www.ce.gunma-u.ac.jp/regpln/katada/kataweb/atop.htm>

- [3] 戸田圭一, 大八木亮, 井上和也, 間畠真嗣, 都市水害時の地下空間の浸水過程について, 京都大学防災研究所年報, 第 47 号 B, 平成 16 年 4 月
- [4] <http://www.mlit.go.jp/river/IDC/index.html>
- [5] <http://www.jwa.or.jp/topics/2005kozui/index.html>
- [6] 谷川, 高橋, 他: Urban Flood Simulation as a component of Integrated earthquake Disaster Simulation System, Proc. 2005 IEEE Int. Workshop on Safety, Security and Rescue Robotics, pp.248-252, 2005,
- [7] <http://www.kedm.bosai.go.jp/>
- [8] 高橋, 谷川, 高橋: 公開地図データにおける建物の自動生成方法, 情報処理学会全国大会 1V-8 2005 3 月
- [9] http://www.pref.aichi.jp/kasen/03_gekitoku/gekitoku_shinkawa/top_geki_sinkawa.html
- [10] <http://www.cbr.mlit.go.jp/kisojyo/rootup/top.html>
- [11] <http://www.pref.aichi.jp/kensetsu-somu/owari-kensetsu/kasen/nakae/GEKITOKU.htm>
- [12] <http://www.city.nagoya.jp/kurashi/shoubou/bousai/kaze/nagoyanokawa/gouu/nagoya00002599.html>
- [13] 渡辺真二, 東海豪雨の観測と解析, 日本気象学会 2001 年度秋季大会シンポジウム
- [14] 片田敏孝, 及川康, 東海豪雨被害における名古屋市天白地区の浸水家庭と銃印の対応に関する実態把握, 河川技術論文集, 第 7 巻, 2001 年 6 月

方策勾配法を用いたサッカーエージェントの学習 ～フリーキックにおけるキッカーとレシーバ～ Learning of Soccer Player Agents Using Policy Gradient Methods ~Kicker and Receiver at Free Kicks~

○中村 浩二 (芝浦工業大学大学院電気電子情報工学専攻)
五十嵐 治一 (芝浦工業大学工学部情報工学科)
石原 聖司 (近畿大学工学部電子情報工学科)

Koji NAKAMURA(SIT), Harukazu IGARASHI(SIT), Seiji ISHIHARA (Kinki Univ.)

Abstract— RoboCup Simulation League is known to be a test bed for research of multi-agent learning. As an example of multi-agent learning in a soccer game, we dealt with a learning problem between a kicker and a receiver when a direct free-kick is awarded just outside the opponent's penalty area. To which point should the kicker kick the ball in such a situation? We propose a function that expresses heuristics on evaluating how the target point is advantageous in sending/receiving a pass safely and contributing to scoring. The evaluation function makes it possible to handle a large space of states consisting of positions of a kicker, a receiver and the opponent's players. A target point of a free kick is selected by the kicker using Boltzmann selection with the evaluation function. Parameters in the function can be learned by a kind of reinforcement learning called policy-gradient method. A point where a receiver should run to receive the ball is simultaneously learned in the same manner. Experiments with four teams that had participated in the past RoboCup competitions showed the effectiveness of our solution.

1. はじめに

近年、人工知能の分野ではマルチエージェント環境における協調行動の学習が研究されている[1]。この研究の題材としてロボットサッカーの競技会である RoboCup が提唱されている[2]。この RoboCup の中の一部門であるシミュレーション部門では、実環境でロボットを動かす難しさから解放されるため、協調的な行動に研究の焦点を当てることができる[3]。一般に、マルチエージェント学習には、状態空間の爆発問題、同時学習問題、不完全知覚問題、実時間処理問題等のマルチエージェント系特有の問題があるとされている[1]。RoboCup サッカーシミュレーションでは、これらの中でも、まず第一に「状態空間の爆発問題」が大きな問題と考えられる。

本研究ではこのように多数のエージェントが広い対象領域に集まり状態空間が大きくなる場合でも適用できる行動学習法の開発を目的としている。今回、その一例としてゴール前でのフリーキックの場面を取り上げた。具体的にはキッカーとレシーバのそれぞれ1名が協調するパス行動において、キッカーのみが学習する場合とキッカーとレシーバが

共に学習する場合の実験を行った。協調行動の学習には強化学習法の一つである方策勾配法と、状態数によらない方策表現とを用いた。

2. フリーキックにおける協調行動

2.1 シミュレーションリーグにおける協調行動

1章で述べた「状態空間の爆発問題」とは、全エージェントの状態数と各状態におけるエージェントのとり行動の組み合わせが多すぎて、理論的には実行可能だが実際問題として時間的に使いものにならなかつたり、メモリオーバーとなつたりする問題である。これは状態空間を適切に設計すれば、ある程度、状態空間の爆発を抑えることができる[1]。

しかし、少領域内で存在するプレーヤーが少数であれば熊田らの研究[4]のようにパス問題を取り扱うことは可能であるが、文献[4]では課題として「フルゲームへ拡張すると計算量爆発を引き起こす」と述べている。つまり、マルチエージェント学習では、仮に小さなシステムにおいて適切な状態空間を設計できたとしても系を大きくすると状態空間の爆発が起きてしまいやすい。したがって、状態空間の表現法を工夫し、状態数によらずにマルチエージェント学習が可能な手法を用いることが望ましいと考えられる。

2.2 先行研究

比較的広い領域で多数のプレーヤーが存在する場合のパス問題において、状態数によらない手法を用いて協調行動の獲得を提案した亀島らの研究例がある[5]。そこでは、まずフィールドを格子状の長方形セルの集合に区切り、passerはルーレット方式でパスを出すセルを決めてパスを出す。もしパスが receiverに通ったら、強化学習でパスを出したセルの価値を高めていき、最終的に一番価値の高いセルにパスを出している。

この手法の特徴は、セルの価値を保存しておくだけでよいので、フィールドにおけるプレーヤーの配置の組み合わせ等といった広大な状態空間を保存する必要がないことである。しかし、この研究では敵プレーヤーの初期位置が固定されていることが前提であり、敵プレーヤーの初期位置が異なると学習結果が利用できず実用性は低い。

2.3 本研究の方針

本研究では、比較的広い領域で多数のプレーヤーが存在する場合のパス問題の一つとして、具体的にゴール前でのフリーキックの場面を取り上げる。そこで、先行研究の欠点を克服するために、行動決定（方策）における状態表現においてプレーヤーの絶対的な位置座標を用いるのではなく、配置から計算されるいくつかの特徴量（ヒューリスティクス）を用いる。

検証実験として、これまでにキッカーのみがパス先の地点を決定するために学習を行い、レシーバはキッカーから指示されたセルへ走りこむといった場面での学習実験を行って、本学習の有効性を確認している[9]。今回は、キッカーだけではなく、レシーバも学習させる。具体的には、キッカーはパス先の地点を、レシーバも走り込む先の地点をそれぞれ独立して両方が学習する実験を行った。

3. 提案手法

3.1 方策勾配法

本研究では強化学習の一種である方策勾配法を用いて学習を行う。方策勾配法とは、報酬の期待値が最大になるように方策パラメータを更新する学習法である。このときの最大化の手段として確率的勾配法を用いる。数学的な基礎がはっきりしており、理論的に取り扱いやすい。また、方策として if-then 型のルールや、ポテンシャルなどの様々な関数が利用できる。方策への知識表現が容易であるという長所がある。元々は、Williams により提案された手法[6]であるが、筆者らも追跡問題やカーリングゲームに適用し、有効性を確認している[7][8]。

3.2 行動決定のための確率的な方策

本稿では、次章で述べるフリーキックの問題のために、以下のようなキッカーの行動決定方式を提案する。まず、一般的なパス行動を考える。フィールドを文献[5]と同様に格子状の長方形セルの集合に区切る。キッカーがセル k へパスを出すという行動 a_k の価値を次のような目的関数で表す。

$$E(a_k; \omega) = -\sum_i \omega_i \cdot U_i(a_k) \quad (1)$$

この関数は、パス先のセル k を決める上で有効と思われる状態の特徴量（ヒューリスティクス） U_i の線形和である。ただし、目的関数 E の値が小さい方が、行動としての価値が高くなるように設計する。

この目的関数を用いて、キッカーは次のボルツマン選択による確率的な方策を用いてパス先のセル k を決定する。

$$\pi(a_k; s) \equiv \frac{e^{-E(a_k; s)/T}}{\sum_x e^{-E(x; s)/T}} \quad (2)$$

ボルツマン選択は温度パラメータ T を大きくするほどランダムに行動を選択するようになり、小さくするほど最も大きい価値の行動を選択しやすくなる。

特に、 $T \rightarrow 0$ では決定論的な行動決定となる。なお、 s は全系の状態（i.e. 全プレーヤー、ボールの位置）を表している。

式(1)では先行研究[5]と違い、敵プレーヤーの数や配置、レシーバの位置など、その場面に応じて各セルへのパスを出す行動の価値を計算しており、プレー中の環境の変化も考慮している。勿論、広大な状態空間の保存を必要としていない。式(1)のパラメータ $\{\omega_i\} (i=1,2,3,4)$ の学習則を次節以下に示す。この学習則を用いてキッカーの方策(2)を学習することができる。

3.3 自律分散的な行動決定とその学習則

一般に、マルチエージェント系全体の状態を s 、行動を a とする。それぞれ、各エージェントの状態と行動とを要素とするベクトルであることに注意する必要がある。行動 a に対する目的関数を $E(a; s, \theta)$ とし、方策が式(2)のようなボルツマン選択である場合に、そのまま方策勾配法を適用すると、パラメータ θ に関する学習則は、

$$\Delta \theta = \varepsilon \cdot r \sum_{t=0}^{L-1} e_{\theta}(t) \quad (3)$$

で表される[7]。 $\varepsilon (>0)$ は学習係数、 r はエピソード終了時に与えられる報酬、 L はエピソード長である。 $e_{\theta}(t)$ は、離散時刻 t における適正度(eligibility)で、次の式で定義されている。

$$e_{\theta}(t) \equiv \frac{\partial}{\partial \theta} \ln \pi(a; s, \theta) \quad (4)$$

次に、行動決定と学習とを自律分散的に行うことを考える。そこで、系全体の方策を、各エージェント i の方策関数 π_i の積で近似する[7]。すなわち、

$$\pi(a; s, \theta) \approx \prod_i \pi_i(a_i; s, \{\theta_{ij}\}) \quad (5)$$

と仮定する。ここで、 a_i はエージェント i の行動であり、 θ_{ij} は π_i に含まれる j 番目のパラメータである。さらに、各 π_i は、各エージェントに定義される目的関数 $E_i(a_i; s, \{\theta_{ij}\})$ を用いたボルツマン分布とする。したがって、3.2の式(2)中の右辺の目的関数 E は、厳密にはエージェントごとに定義された目的関数 E_i である。

なお、(5)の近似は、エージェント間の行動選択の相関を無視していることに相当している。また、他のプレーヤーの位置情報も分かっているという前提に立っている。各 π_i には系全体の状態 s が使用されている。したがって、各エージェントは味方や敵の行動選択とは無関係に行動を選択するが、彼らの存在（位置情報）までを無視しているわけではないので、このような近似を行っても協調行動を学習することはある程度までは可能と考えられる。

一方、報酬 r の期待値 $E[r]$ をパラメータについて微分すると、形式的には、

$$\frac{\partial E[r]}{\partial \theta_{ij}} = E \left[r \sum_{t=0}^{L-1} e_{\theta_{ij}}(t) \right] \quad (6)$$

となる。(5)の近似を用いて(6)の右辺の適正度を計算すると、

$$e_{\theta_{ij}}(t) \equiv \frac{\partial}{\partial \theta_{ij}} \ln \pi(a_i; s) \quad (7)$$

$$\approx \sum_i \frac{\partial}{\partial \theta_{ij}} \ln \pi_i(a_i; s, \theta_{ij}) \quad (8)$$

さらに、各エージェントの目的関数中のパラメータは互いに独立であると仮定すると(7),(8)は、

$$e_{\theta_{ij}}(t) \approx -\frac{1}{T} \left[\frac{\partial E_i}{\partial \theta_{ij}} - \left\langle \frac{\partial E_i}{\partial \theta_{ij}} \right\rangle \right] \quad (9)$$

と表される。一方、パラメータ θ_{ij} の学習則は(3)の導出と同様にして、(6)を用いて確率的勾配法を適用すると、

$$\Delta \theta_{ij} = \varepsilon \cdot r \sum_{t=0}^{L-1} e_{\theta_{ij}}(t) \quad (10)$$

となる。ただし、(5)の近似により(10)の右辺の適正度は(9)で与えられる。(3)が系全体の目的関数 $E(a; s)$ に含まれるパラメータに関する学習則であるのに対し、(10)はエージェント i が自己の行動 a_i を選択するために使用する目的関数 E_i の中に含まれるパラメータの学習則であり、かつ、自己の目的関数 E_i だけを用いている点で、自律分散的な学習であると言える。

4. 問題設定と学習実験

4.1 プレーヤーの配置

実験では図1に示すようにペナルティエリア付近を一辺5mの正方形セルで4×8に分割したフィールドを用いる。さらに均等にゴールも3分割する。これにより、「ゴールにパスをする」ことが「直接シュートをする」ことを実現している。

使用するエージェントと配置場所を以下に、その配置例を図1に示す。なお、**SoccerServer**ではフィールドの中心(センターマーク)を原点とし、タッチラインに平行にX軸、ゴールラインに平行にY軸をとる。プレーヤーは4種類で、キッカー1人、レシーバ1人、敵チームのディフェンダー1人とゴールキーパー1人である。

キッカーにはフリーキック時にボールを蹴るだけの役割を与える。配置場所のX座標はゴールラインから24mの位置で固定、Y座標はランダムな位置とする。レシーバはキッカーが指示したセルに走りこみ、パスを受けたらその場で直ちにシュートをする。配置場所はオフサイドポジションにならないランダムな位置とする。

敵ディフェンダーはゴールを決めさせないためにパスカットやプレスを行う。配置場所はゴール3セ

ルを除くセルの中でランダムな位置とする。ゴールキーパーはゴールエリア内のランダムな位置に配置する。

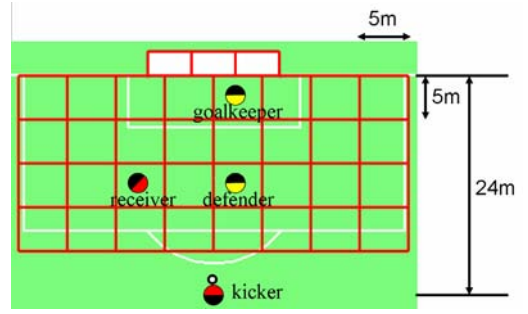


Fig. 1 Arrangement of players

プレーヤーの他に **trainer** というエージェントも使用する。**trainer** はプレイモードを **kick_off** から **free_kick** に変更したり、プレーヤーとボールを移動したりすることができる。そして、フィールドを監視し、パスは通ったのか、シュートは決まったのかをキッカーに伝えることができる。

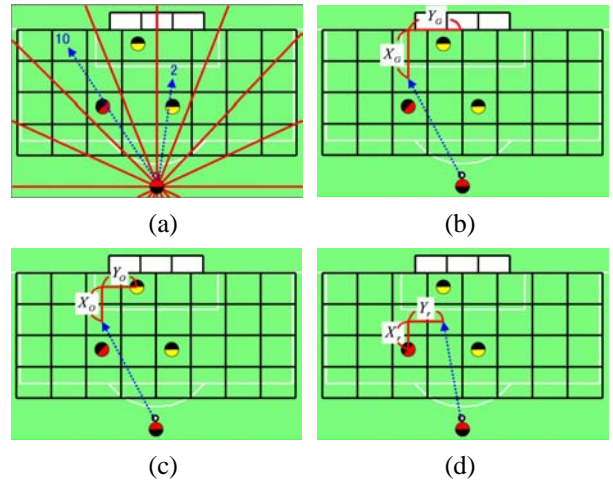


Fig. 2 Heuristics $\{U_j\} (j=1,2,3,4)$ used in $E_i(a_k; s, \{\theta_{ij}\})$

4.2 目的関数に用いたヒューリスティクス

実験ではゴール前でのフリーキックにおいて有効と思われる以下の1)~4)の4つのヒューリスティクスを用いた。なお、どの関数も値が大きいほど価値がある。

1) パスコースにおける敵の有無：

$$U_1 = \begin{cases} 10.0 & \text{敵がいない場合} \\ 2.0 & \text{敵がいる場合} \end{cases} \quad (11)$$

実際にパスをするにあたって、パスコースに敵がいない方がパスは通りやすい。図2(a)のようにキッカーの正面180度を22.5度ずつ8領域に分割し、敵がいれば2、いなければ10を返す。

2) パス先とゴールとの距離：

$$U_2 = -(X_G + Y_G) / 3.5 \quad (12)$$

パス先とゴールとの距離に近い方が、パスを受け取った時にゴールに結びつきやすい。式(12)の X_G と Y_G は図 2 (b)のようにそれぞれパス先とゴールとの距離の X 成分, Y 成分である。なお, 式(12)は $0.5 \leq U_2 \leq 10.0$ と正規化してある。

3) パス先と最近接の敵との距離:

$$U_3 = (X_o + Y_o) / 5.0 \quad (13)$$

パス先と敵との距離が遠いほど, そのパス先にはスペースがあるということである。つまり, フリーでシュートを打ちやすくなる。式(13)の X_o と Y_o は図 2 (c)のようにそれぞれパス先と最近接の敵との距離の X 成分, Y 成分である。なお, 式(13)は $0.0 \leq U_3 \leq 10.0$ と正規化してある。

4) パス先とレシーバとの距離:

$$U_4 = -(X_r + Y_r) + 50.0 / 5.0 \quad (14)$$

パス先とレシーバとの距離に近いほど, パスは通りやすくなる。式(14)の X_r と Y_r は図 2 (d)のようにそれぞれパス先とレシーバとの距離の X 成分, Y 成分である。なお, 式(14)は $0.0 \leq U_4 \leq 10.0$ と正規化してある。

4.3 報酬 r

報酬 r は 1 エピソード終了後にキッカーとレシーバに以下のように与える。

- パス失敗時 -0.5
- パス成功∧シュート失敗時 3.0
- パス成功∧シュート成功時 10.0

ただし, 1 エピソードを 70 シミュレーションサイクル(7sec)とし, 1 エピソード終了前にゴールを決めることができれば, エピソードはそこで終了とする。

4.4 キッカーとレシーバの学習実験

2つのエージェントが独立に学習を行っても, 共通の方策を獲得する「合意形成」[5]が 4 章で述べた本方式の枠組みで可能であることを検証するため, 以下の環境下でキッカーとレシーバの同時学習を行った。まず, キッカー1人, レシーバ1人, ディフェンダー1人, ゴールキーパー1人の2対2の対戦とする。次に, 使用する敵エージェントは Trilearn Base である。

Trilearn Base とは 2003 年に優勝したアムステルダム大学チーム UvA Trilearn 2003 から高度な行動決定や戦略を除いてソースコード形式で配布されているチームである。したがって, UvA Trilearn 2003 よりは弱くなっているが最低限の行動は保証されている。

上記の準備の下で, 学習を次のような流れで 1000 エピソード行った:

- ① trainer がプレイヤーを適切に配置
- ② キッカーとレシーバはそれぞれ目的関数を計算し, パス先または走りこむ先セルを独自に決定
- ③ キッカーはパスを出し, レシーバは走りこむ
- ④ パスを受けたらその場でシュート

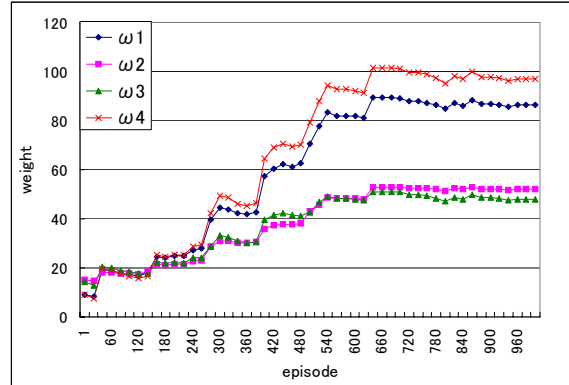
⑤ 結果に対して報酬を与え, 重みを更新
 なお, ④でレシーバがパスを受けたその場ですぐにシュートをしているのは, 本手法ではパス行動だけを評価したいので, ドリブルなどを行うとパス先がよかったのかが分からなくなるからである。

5. 実験結果と考察

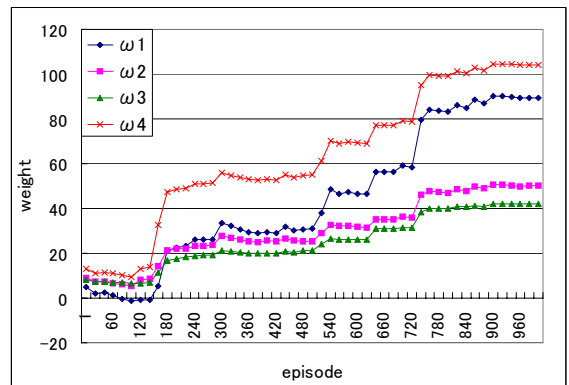
4.で述べた条件の下で, Trilearn Base を相手にしてキッカーとレシーバを独立に学習を行った。重み $\{\omega_i\}$ の初期値はランダムとし, 学習中の変化を図 3 に示す。パス成功率と得点成功率の結果を図 4 に, 報酬の期待値を図 5 に示す。なお, 2つの成功率の値はその学習回数における重みの値を用いて 100 エピソードを 10 セット行った平均値である。

図 3 より 1000 エピソードの学習における重みの初期値が異なっても, キッカーとレシーバの重みの値は学習が進むに連れてほぼ一致していく。これによりキッカーとレシーバが共通の知識(方策)を獲得できたことがわかる。

図 4 よりパス成功率は 36.2 ポイント, 得点成功率は 12.1 ポイント上昇した。どちらも学習前の 2 倍以上の値である。しかし, パス成功率の大幅な上昇のわりに得点成功率が伸びていないのは, 図 3 より「パス先とゴールとの距離」の重みが小さく「パス先とレシーバとの距離」の重みが大きいからである。つまり, パス先を決定する上でレシーバとの距離が優先されるので, パスはよりたくさん通るようになる。



(a) Kicker



(b) Receiver

Fig. 3 Change of weight

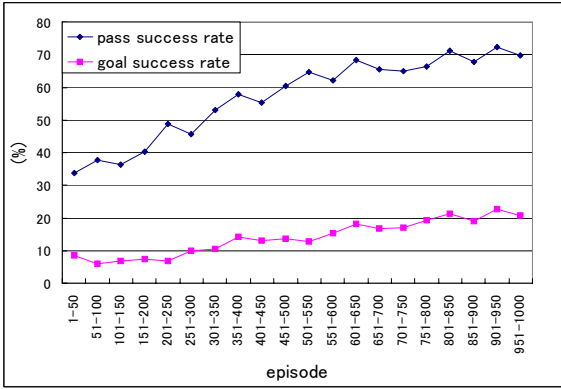


Fig. 4 Pass/goal success rate

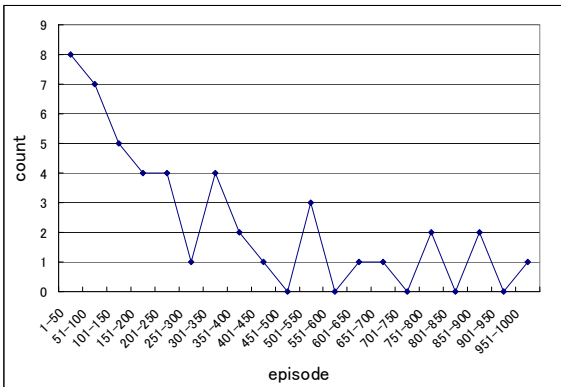


Fig. 5 Number of direct shots

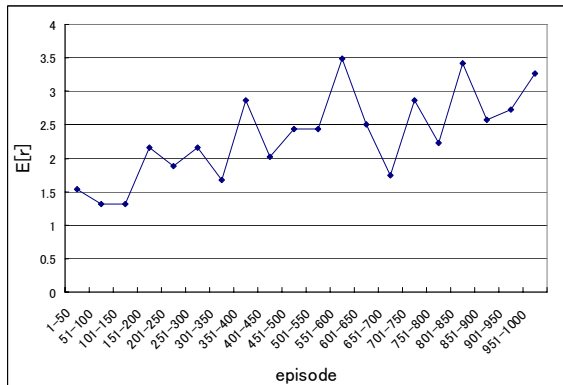


Fig. 6 Expected value of reward

一方、ゴールとの距離はあまり優先されなくなるので、ゴールから遠いところでボールをもらうことがたくさんあり、得点成功率はそれほど上がらない。それでも得点成功率が改善されたのは、ゴール付近へのパスでパスが通れば得点、通らなければキーパーに取られるといったギャンブル性のパスよりも、ゴールから遠くてもたくさんパスを通してたくさんシュートを打ったほうがより得点を奪うことができることをエージェントが選択したと思われる。

また、学習の初期段階 (50~150) で得点成功率が一度下がりまた上昇するといった現象が見受けられた。図6の報酬の期待値の変化からも同様なことが見て取れる。これは重みの初期値がランダムであるが故に、図5の直接シュート数の変化からわかると

おり初めは直接シュートを狙うことが比較的多く、シュート数が増えることにより得点を取ることができたためである。その後、協調的なパス行動を探索するため一時的得点成功率が下がる (報酬の期待値も下がる) 段階を経た後、パス成功率の上昇と共に得点成功率が上がっていく過程を踏んでいるのが見て取れる。

6. 行動選択の相互作用についての考察

6.1 相手の行動に依存した行動選択

本章では、各エージェントが行動を決定する際に生ずる相互作用について考察したい。3.3の式(5)では、各エージェントの行動決定に際しては、味方の行動決定とは無関係に決定するという近似を行った。しかし、味方の状態 (位置情報) と無関係に行動決定を行ったわけではないので、学習により協調行動が発現する可能性がある。実際に、式(13)や式(14)に示されているように、敵や味方の位置情報がキッカーの出すパス先の地点とレシーバの走り込む先の地点との決定に大きな影響を与えている。このことが、5章の実験で示されたようなキッカーとレシーバとの合意形成のための原動力になったと考えられる。

しかしながら、4章で述べた学習方式では、学習過程において確率的な方策を用いていることもあり、重みの初期値の設定によっては必ずしも合意形成に至るとは限らない。行動決定に必要な情報の種類と、目的関数、また報酬関数についても同じものを使用しているため、たまたま重み係数の比率が同じ値に収束していったと解釈することもできる。実際に、キッカーとレシーバとが行動選択で用いた目的関数の中には、両者の行動を一致させるような項は用いられていない。逆にもし、このような項が存在すれば、たとえ両者が利用できる情報 (たとえば視覚情報) や知識 (目的関数) が完全に一致していなくても、合意形成を促進するのではないかと考えられる。

そこで、ひとつの方法として、目的関数に相手の行動を考慮した項 (相互作用の項) を加えることが考えられる。たとえば、キッカーとレシーバとがそれぞれの行動 a_K, a_R を選択するために使用する目的関数を次の $E_K(a_K), E_R(a_R)$ で与える。

$$E_K(a_K; \omega) = -\sum_{i=1}^4 \omega_i U_i(a_K) - \omega_5 U_5(a_K, a_R) \quad (15)$$

$$E_R(a_R; \omega') = -\sum_{i=1}^4 \omega'_i U_i(a_R) - \omega'_5 U_5(a_K, a_R) \quad (16)$$

ただし、(15),(16)中の $U_1 \sim U_4$ は、4章で定義した関数であり、 $U_5(a_K, a_R)$ は次のように定義する。

$$U_5(a_K, a_R) \equiv -(X_{KR} + Y_{KR}) / 50.0 \quad (17)$$

式中の X_{KR}, Y_{KR} はキッカーのパス先のセルとレシー

バが走り込む先のセルとの間のベクトルの X 成分, Y 成分である. この $U_5(a_K, a_R)$ の項は, キッカーのパス先とレシーバの走り込む先とを近づける働きをする.

6.2 相手の行動の観測または予測

前節で述べた相互作用の項 U_5 を計算するには, 互いに相手が選択した行動を知る必要がある. また, 同時に行動選択をするとデッドロックに陥るなどの不都合な事態を生ずる恐れも考えられるので行動選択の順序も重要である. 本節ではこの2つの点について考察する.

まず, 相手が実際に選択した行動を知るには, 自らの観測による方法と相手との情報交換などの通信による方法との2つが考えられる. 前者の場合は相手の行動実行後に行動選択を行う必要がある. また, 観測や通信が可能であるという前提も当然必要である.

次に, 実際に観測や通信ができない場合や, 行動選択の順序が相手よりも先でないといけない場合には, 相手の行動を予測する必要がある. そのときには, 例えばキッカーがレシーバの行動を予測するには, (16)の U_5 を除いた $U_1 \sim U_4$ の項だけからなる目的関数 $E_R(a_R)$ の最小値を与える行動 a_R を予測値として使用する(あるいはボルツマン選択のように確率的に選ぶ)という方法が考えられる. すなわち, 相手行動を予測するには, 相互作用の項を除いた相手の行動選択モデルを用意して利用するという予測方式である. この予測の精度を高めるには, 各々の行動モデルである目的関数中の重み係数の値を時々交換するなどの処理が必要であろう. またこうした予測も, 各エージェントが別々に独立したタイミングで行うのか, 一斉に行うのか, あるいは何らかの優先順位が設定されているのかと言った種々の場合が考えられる. 対象とする問題や状況に応じて, 観測, 通信, 予測のための方式を工夫する必要がある.

7. まとめ

本研究では, マルチエージェント系における行動学習の一例として RoboCup シミュレーションリーグにおけるゴール前のフリーキックの場面に適用し, キッカーとレシーバの双方が学習する問題を取り上げた. また, 「状態空間の爆発」を抑えるため, 状態数によらない方策表現と方策勾配法とを用いた協調行動獲得法を提案した. その結果, パス成功率と得点成功率が学習前と比較して2倍以上となり, 提案手法による学習が有効であることを確認することができた. さらにキッカーとレシーバが共通の知識(方策)を獲得でき, 提案手法の枠組みで合意形成が行われることを確認した. なお, 本実験では学習するエージェントはキッカーとレシーバの2名だけであったが, 学習が自律分散的であるため, 学習するエージェントの数を増やしても学習時間が爆発的に増えることはない.

今後の予定としては, キッカーとレシーバとでヒ

ューリスティクスが異なる場合や, 得られる視覚情報が不完全で互いに異なる場合にどれだけ学習能力があるのかを調べて行きたい.

また, 各エージェントがパス行動とレシーブ行動との2つの方策を切り替えることにより, 複数プレイヤー間での複数回のパス交換を実現することを考えている. さらには, 通常の試合中での一般的な協調行動の学習へと研究を進めて行く予定である.

謝辞: 本研究の一部は日本学術振興会科学研究補助金, 基盤研究(C) (課題番号 15500125) を用いて行われた.

参考文献

- [1] 高玉圭樹: マルチエージェント学習-相互作用の謎に迫る-, コロナ社, (2003)
- [2] RoboCup Official Site <<http://www.robocup.org/>>
- [3] 野田五十樹: シミュレーションリーグとインフラ技術の技術的課題と展望, 日本ロボット学会誌, Vol.20, No.1, pp.7-10, (2002)
- [4] 熊田陽一郎, 植田一博: 予測能力を持つサッカーエージェントによる協調戦術の獲得, 人工知能学会論文誌, Vol.16, No.1, pp.120-127, (2001)
- [5] 亀島力, 遠藤聡志, 山田孝治: 強化学習を用いた共同注視点に基づく合意形成の獲得, 電子情報通信学会・信学技報, Vol.99, No.131, pp.29-35, (1999)
- [6] Williams, R.J.: Simple Statistical Gradient- Following Algorithms for Connectionist Reinforcement Learning, Machine Learning, Vol.8, pp.229-256, (1992)
- [7] 石原聖司, 五十嵐治一: マルチエージェント系における行動学習への方策勾配法の適用-追跡問題-, 電子情報通信学会論文誌 D-I, Vol.J87-D-I, No.3, pp.390-397, (2004)
- [8] 五十嵐治一, 石原聖司, 野原 勉: 方策勾配法を用いた運動方程式中のパラメータ学習-2ストーン系のカーリングゲーム-, ロボティクス・メカトロニクス講演会'05(ROBOMEC'05)講演論文集, 1A1-N-028(pp.1-4), (2005.6.10-11, 神戸, 主催: 日本機械学会ロボティクス・メカトロニクス部門)
- [9] 中村浩二, 五十嵐治一, 石原聖司: 方策勾配法を用いたフリーキック時の行動学習, ロボティクス・メカトロニクス講演会'05(ROBOMEC'05)講演論文集, 2P1-S-022(pp.1-4), (2005.6.10-11, 神戸, 主催: 日本機械学会ロボティクス・メカトロニクス部門).

CITIZEN Eco-Be! League: bringing new flexibility for research and education to RoboCup

Rodrigo da S Guerra¹, Joschka Boedecker¹,
Norbert Mayer^{1,2}, Shinzo Yanagimachi³,
Yasuji Hirose³, Kazuhiko Yoshikawa³,
Masaaki Namekawa³, and Minoru Asada^{1,2}

¹Dept. of Adaptive Machine Systems,
²HANDAI Frontier Research Center,
Graduate School of Engineering, Osaka University, Osaka, Japan
³CITIZEN Co., Tokyo, Japan

Abstract

The RoboCup 4-legged and Simulation leagues offer good examples of the benefits that can be achieved by using a standardized platform that allows for an easy comparison of results and concepts in robotics research. However, there is a lack of a standard platform that provides the flexibility of simulations in real robots at a reasonable price. To address this, we introduce a new setup based on a project sponsored by CITIZEN Co. Japan. It allows for a huge variety of different usages for a large number of mini-robots in an augmented reality (AR) environment at quite inexpensive costs. The core of the project is the CITIZEN Eco-Be! mini-robot. It is built with cutting edge technology in miniature robotics using high quality and low-cost CITIZEN watch technology. Finally, a new RoboCup league is proposed which should allow RoboCuppers to face an assortment of new interesting and challenging research issues, all based on the presented setup.

1 Introduction

Generally speaking, doing research on robotics is an expensive task, and the RoboCup real-robot leagues are no exception. Even in the most inexpensive real-robot league participants would spend, on average, several thousands of dollars in order to have a team. As for the main reasons, one could surely account for the unavailability of commercial platforms adequate for the games, thus bringing the need for custom brew robots. Moreover, the strong competition force teams to challenge themselves to come up with new design ideas which quite often are translated into more complex and expensive hardware. This last factor also implies that a wider spectrum of fields need to be covered for the complete design of the machines, including technical issues which are not always related to the research focus originally in mind. Such difficulties may seem inherent to the research track

of some institutions, but they also prevent the participation of those who do not have the man-hour and the money for the journey.

Another common problem in robotics which is frequently overlooked is the lack of a standard platform in which one could easily compare results of different methodologies. The two-dimensional environment of simulation league has already proven for years the advantage of having such standardization: papers often show comparative results using the standard environment and playing against a provided adversary (usually the code of a good team of the former years). To a minor degree the four-legged league also share some of these advantages – for instance, several teams use the same low level control skills for controlling the gait of the robots.

CITIZEN Co. Japan, one of the biggest sponsors of the RoboCup initiative [Kitano and Asada, 2000], is a renowned company, especially in regard to miniature devices. Currently the company is committed to the endeavor of developing a miniature-yet-affordable robotics platform. This comes in opposition to the main stream in robotics, which generally points toward costly solutions. Instead, CITIZEN Co. Japan focused on versatility and affordability, taking advantage of their well-established watch technology for the development of inexpensive miniature robots.

A new research project in Osaka University is about to bring new flexibility in an endless number to possible applications with the introduction of an augmented reality (AR) environment to be used along with this new robot. Huge improvements in versatility, flexibility, and standardization are among the merits of such an AR environment.

This document proposes a new league (tentative name "Mini League") in which Eco-Be! would be used in a standard and flexible environment for a variety of competitions in a mixture of reality and simulation. The presented AR setup brings new challenges in a variety of areas which RoboCup has not yet fully involved so far. Based on this we propose the conceptualization of a new RoboCup league. This paper is composed of three major parts: 1) Introduction of the new mini-robot currently



Figure 1: An Eco-Be! prototype. In this robot the differential wheels are driven by two motors which are battery-powered. Besides the PIC controller all other hardware comes from devices originally designed for CITIZEN watches.

under development by CITIZEN Co. Japan; 2) Description of the AR environment for these robots which is being developed in a joint project between CITIZEN Co. Japan and Osaka University.; and 3) Proposal for a new RoboCup league based on this new architecture.

2 Eco-Be! The Robot

The Eco-be! is a robot of very small size (approximately 1.6 x 1.6 x 2.1 cm) which can be produced at relatively low costs by using CITIZEN's renowned knowhow in miniature devices and their well established watch technology [Yoshikawa, 2002]. This allows the employment of large number of robots in a rather reduced space with a very low budget and amazing portability. For an illustration of the early prototypes see figure 1.

The first RoboCup prototypes were presented to the public in March 2006. These first prototypes include sophisticated control over the differential wheels through an IR interface, and they are battery powered in order to assure longevity and necessary performance of the motors. In Bremen even more sophisticated robots shall appear for demonstrations, this time with considerable improvements in the fine-grained controlling of trajectories and velocities. Following this pace, CITIZEN proposes long term development of increasingly more sophisticated Eco-Be! robots over the years. There is plenty of room for improvement in what regards to modern watch technology, including miniature devices for bluetooth, camera, etc. – the robots would be updated accordingly to the needs observed during the years. Furthermore, these robots may become part of CITIZEN's line of products, provided sufficient market acceptance and robust development becomes apparent over the years.

3 Eco-Be! The AR Environment

The figure 2 illustrates the suggested setup (see caption for details). The idea for this setup came from [Kazama *et al.*, 2005]. The software architecture would be borrowed from the current three-dimensional simulation league, thus ensuring the agents would be individually programmed as isolated agents. Such setup brings

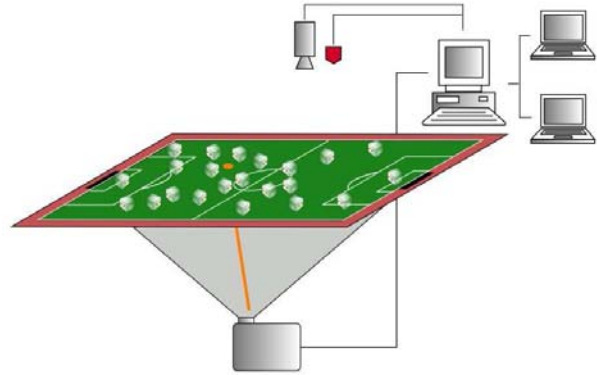


Figure 2: Setup for the Eco-Be! AR environment. The several robots follow commands given by a central server through an IR beam. The actual position and orientation of the agents is feedback to the server by a camera located on the top. The environment along with its features is projected onto the field by using a multimedia projector or any sort of high definition TV screen installed below. Optionally the IR sensor on the top could be substituted by a color sensor on the bottom, allowing more localized control of each individual by varying beams of color under each robot.

huge flexibility for the simulation of sensors and actuators not yet implemented in the real robots. This means the mixture of reality and simulation allows for the use of robots years ahead in design despite their current limitations. At the same time this AR setup may guide the development of new devices to be equipped in robots during the years, gradually migrating towards fully stand-alone robotic agents – which could still be optimized by AR (there is always room for improvement).

In general this setup is also similar to the one used in the small size (F180) league [Sma, 2006] or the ULeague as proposed in [Anderson *et al.*, 2003]. However, the AR environment as described here allows for a lot more flexibility, e.g., in different kinds of experiments that can be performed conveniently. By using the suggested AR setup, projections of environmental features surrounding the robots can be used instead of real objects or arenas. Among the main advantages of using this AR environment, one could mention:

- Great flexibility on the design of the task due to the control of the parameters of virtual objects/features (e.g. ball diameter and its dynamics can be easily changed);
- Possibility for an enormous variety of creative applications that include, but are not limited to, soccer. Among the numerous conceivable applications there would be some which are almost (or completely) impractical to do with real robots if not for the aid of augmented reality (e.g. simulation of force fields, artificial webs, laser beams);
- Environment becomes more visually appealing, turning it interesting for the audience and for the

media. This also allows the display of a variety of usually unseen features, making it also appealing for applications on a variety of new fields such as swarm intelligence (e.g. bug heat fields or ant trails can be shown) and artificial life (food, preys and predators can be simulated in big number and any shape).

- Robots would not be required to be equipped with extra sensors and could still enjoy complex models of the environment through the use of virtual sensors and actuators (e.g. virtual camera, heat sensors, kicking device, virtual grippers, etc.). Controversial enough, this removes the need for expensive/complex equipment to be installed in the robots but actually extends the possibilities to the limit of imagination.
- Standardization of both software and hardware, making it possible to do benchmarks of different techniques used by different people in a straightforward way.
- Portability, for the case of using a multimedia projector – 22 robots plus a projector could easily fit into a briefcase.

4 Eco-Be! The Mini League

This document proposes a new league (tentative name is "Mini League") using the Eco-Be! robots¹. Competitions would be held initially in an autonomous soccer environment similar to that of simulation league. A variety of competitions and challenges would be gradually included over the years covering a diverse range of fields in reflex to the general interest and quorum of development. Throughout this section several possible competitions will be illustrated.

4.1 Mini League Autonomous Soccer Games

This is actually the first ongoing development application of the Eco-Be! robots for use within RoboCup. It would consist in an augmented reality realization of 3D Simulation League games. A total of 22 Eco-Be!s would be disposed of a small arena on which the virtual field, including a virtual ball, would be projected instead of using an actual kicking device and real ball. In this way the games would be made interesting for the spectators. For implementing this, the RoboCup Simulation League architecture is being taken as a starting point. The current robot soccer simulator is designed with plugin capabilities that allow extensions to be added with certain ease. Simulation League games are the only ones with 22 agents involved so far and they feature highly developed strategies and teamplay. However, for spectators, it is sometimes too difficult to understand what is happening in the simulators. Also, the current visualizations sometimes lack attractive visual features. Using real mini robots like the Eco-Be! for this purpose avoids those shortcomings of the simulation. Given the small size of the robots, they probably have a "cuteness factor" that makes them fun to watch if they engage in these kinds of soccer games.

¹<http://er04.ams.eng.osaka-u.ac.jp/ecobe-robocup/>

4.2 Mini League Interactive Soccer Games

The proposal above could be extended to also allow interaction of spectators with the ongoing game. While it is fun to watch the robots playing soccer, it should be even more exciting to be able to actively take part in the game either controlling some of the players individually, or even controlling a whole team at once by interacting with its strategy online during the game. Kids are very good at using game pad controllers in computer games, so those would be the devices of choice for them to control individual robots. In contrast, here more skilled adults could interact with the team formation and player roles by sketching strategies directly on a screen of a tablet PC. The AR field could be easily shown at the screen of such a PC and the human could intuitively interact with the team by the use of pen strokes. Team strategy diagrams could be drawn in a way similar to the ones human coaches use in soccer and basketball games. Competition could be additionally held on the development of the most intuitive method for recognizing the intended orders of an untrained spectators. This would foster research towards more friendly and intuitive human/machine interfaces.

The issues that would be interesting for research in the proposed setup would include:

- **Human-computer interaction issues**

See for instance the research of professor Nishino in the OZRP project [Nishino, 2005].

- **Strategy acquisition**

The agents in the users team should analyze the movements of the human controlled robots and try to learn from successful moves. An example of this (with logfiles instead of online, user controlled movements) can be found in [Nii *et al.*, 2005]).

- **Teammate and opponent modeling**

The agents of the team including the human players could try to adapt to their human partners by learning and using a teammate model. The autonomous agents of the other team could, on the other hand, try to build an opponent model of the team including the human players and adapt their strategy accordingly (see for instance [Riley and Veloso, 2001]).

- **Adaptive skill learning**

Based on the strategy and the control of the human players, the autonomous agents could try to adapt their parameters according to their observations during the game (see [Kawarabayashi *et al.*, 2005] for an example).

4.3 Mini League Rescue

Robot navigation and obstacle avoidance are among the older fields of research on autonomous agents. More recently, topics such as sensor fusion and intelligent (adaptive/learning) behavior became also topic of prominent research. In near future, autonomous home robots have to be able to merge all the information of the various sensors for successfully navigating indoors, avoiding both fixed obstacles (walls, doors, staircases), and mobile obstacles (furniture, people). Other issues include limited

battery power and eventual failure of small subsystems. Robots have to deal with all these issues and still be able to effectively conclude their goal task with satisfactory performance and robustness. The arrival of the first personal robots in the near future points toward the strong importance of allowing the employment of such home robots for indoor first aid in case of disasters.

The Eco-Be! robots are two wheeled, differential driven robots, which are the most popular configurations for research autonomous wheeled robots. The differentially driven axis adds a restriction constraining movements in the direction of the axis of the wheels – the so-called *non-holonomic restriction* (see for instance [Kühne and Walter F. Lages, 2005]). Most of the practical real-world vehicles are constrained by non-holonomic restrictions and many other applications such as robot manipulators also share non-holonomic restrictions. The control of non-holonomic vehicles is not straight forward because the relation of the rotation of the wheels to the actual displacement of the robot is not linear. Navigation of non-holonomic vehicles in clustered environments is a challenging topic of research, and when talking about rescue robots, then performance is a must.

An indoor environment can be topologically represented by projecting features into the field. The AR field allows for realistically animated representation of features such as:

- Floor, walls, doors, windows, staircases. The robot could be equipped with simulated actuators (e.g. grippers) for opening doors, or simulated "steering wheels" for going upstairs. The whole scenario can be easily changed to show a different floor level.
- Mobile furniture such as chairs, tables, TV, shelves, etc. The furniture can be represented with realistic physics, allowing the robot to eventually drag some of the objects in order to gain access to some area.
- People, both walking and lying down. Scared people running around, or even injured unconscious people laying around waiting for help can be simulated.
- Explosions, fire, water, or even earthquakes can be realistically simulated. Because besides the robots, all other objects are simulated, they can catch on fire, or even walls can be destroyed by explosions.

The Eco-Be! robots can be equipped with any sort of local simulated sensors, among which one could include: fire or heat detectors, cameras, resonance distance sensors, compass, GPS, range finders, microphone, etc. These sensors can be simulated in a way such as to avoid the burden of actually designing sensor controllers for parsing raw data, thus allowing teams to concentrate on more high level strategic planning and navigation issues instead. For example, in a realistic scenario, a camera may return a label identifying the object seen with more certainty than a laser range finder can do, but the later has more accuracy on measuring distances. Additionally, microphones can automatically "recognize speech", and so on. This allows the research on a realistic but years beyond in a futuristic scenario. Obviously noise and eventual failure of individual subsystems should be

simulated as well.

4.4 Mini League Emergency Traffic Control

In a more macro level, the traffic of big urban centres is an issue when talking about big disasters such as typhoons, earthquakes and tsunamis. At the same time, the traffic flow should allow constant evacuation of vehicles from dangerous areas at a reasonable rate, but should also allow the rescue team to reach the rescue spots as quickly as possible.

Considering such a scenario, a small portion of a city can be represented in the field. The simulated city would be represented by topological roads scaled in such a way for a number of Eco-Be! robots to be the vehicles. A small quantity of Eco-Be!s would represent rescue vehicles while the other big majority (maybe dozens) would represent the ordinary traffic. In contrast to the previously mentioned application, here the vehicles are not directly steered by the teams. The teams should rather concentrate their development on the elaboration of adequate traffic control by using traffic lights and simulated human agents that can go onboard some rescue vehicle. These rescue vehicles would drive into some region, deploy people for aiding in the traffic control on demand and also deploy people for first-aid, firefighting, etc. Additionally, the teams would have the task to plan the quickest route for rescue vehicles to get into some target position, and also decide what target positions to reach (priority, timing, etc.) Remaining ordinary vehicles would be centrally controlled, with inclusion of simulated stress factors, such as traffic accidents, erratic behavior, and hysterical runaway of some vehicles. According to their role (police, fire truck, etc.), the Eco-Be!'s could be colored either by using paint or through projection of the respective colors onto the robots. The movements of the robots in the virtual environment should then be much easier to track (and more fun to watch) than it is possible now on the big screens.

4.5 Mini League AI

Swarm intelligence and Artificial Life are perhaps the most specialized applications in which Eco-Be! can fit amazingly well and most other platforms can not. As for the reasons, there is:

- Low-cost hardware – Eco-Be! can probably be produced at a very low cost, and even lower cost if mass produced by CITIZEN. This means a researcher should be able to buy, let's say, a hundred of Eco-Be! robots for research for less than one thousand US dollars.
- Projective Augmented Reality adds flexibility – Multimedia Projectors are not as expensive as they used to be. Video cameras, capture cards and PC computers are also getting more and more affordable as years pass. This means the scientist has access to an affordable platform for controlling a huge number of robots in a virtually uncountable number of swarm intelligence [Bonabeau *et al.*, 1999] or artificial life [Langton, 1997] experiments. Moreover,

while for soccer games or rescue operations the augmented reality has some appealing factor, in swarm experiments the flexibility of the augmented environment becomes a real aid for visualizing usually unseen factors such as heat, pheromones, etc.

- Small size – Eco-Be! robots are small enough for allowing a huge quantity of robots to fit into the small rectangle defined by the projected environment. Because the robots are so small and the projector can be turned off, this means the whole system can be easily transported to a different place and hundreds of robots carried into a single briefcase.

Competitions could be held on the application classic swarm intelligence problems such as *ant-like pheromone trailing*, *heat seeking bugs*, and artificial life subjects such as *dominance and role assignment*, *foraging*, *genetic evolution*, *prey/predator equilibria*, etc. Applications range from very theoretical computational problems (e.g. traveling salesman) to very specific biological models for evolution, disease dissemination to cite a few.

4.6 Mini League Education

The Eco-Be! robots shall be produced for a relatively low price which makes it an attractive platform for schools and research institutions with smaller budgets. Also, the scalable complexity that augmented reality brings to the robots makes them suitable as a didactic tool for education in a variety of undergraduate subjects. Thus the Eco-Be! robots could be used to bridge the currently existing gap between the RoboCup Junior league (which is for high school students), and the more advanced leagues in which mostly master and PhD course students are involved.

Among the subjects of undergraduate that could be faced we could cite:

- Programming (both structured and object-oriented);
- Control Theory (feedback control, non-holonomic systems);
- Multiagent programming (non-Markovian processes, cooperation, competition);
- Foraging and Ecosystems (ecologic equilibria, dominance);
- Artificial-life and genetic algorithms (evolution and adaptation to environment);
- Swarm intelligence (ants pheromones, flocks behavior, etc.);
- Reinforcement Learning (grid world, sarsa, tile coding, etc.);
- Computer Graphics and Augmented Reality
- Human-Machine interface;

The forementioned augmented reality architecture and a few dozen of Eco-Be! robots should be affordable and the setup reproducible by the majority of colleges, universities and high level technical schools around the world. Eco-Be!'s AR setup could span all levels of education from elementary school teaching to graduate research, provided the adequate application design. If adequately exploited, such application has enough appeal

to become an standard platform for education and research on some/all of above listed fields. The advantage of having a standard platform and shared development of applications is enormous. Just for instance, this could provide an standard platform allowing results to be easily compared and reproduced among different institutions during conferences and workshops (reproducing and comparing results is usually a big problem in robotic research).

5 Discussion

This paper presented the Eco-Be! robot and described an AR environment. This architecture was then proposed as a standard for making a new versatile and affordable real-robot RoboCup league, here called “Mini League”. As already mentioned, development has already begun on the robots themselves, on the AR setup and on the first demonstration of autonomous soccer application using the whole environment. The entire architecture is supposed to be ready for being displayed already in Bremen this year. If well accepted by RoboCup Federation the new league shall be ready for its first steps during RoboCup of 2007.

References

- [Anderson *et al.*, 2003] John Anderson, Jacky Baltes, David Livingston, Elizabeth Sklar, and Jonah Tower. Toward an undergraduate league for robocup. In *Proceedings of Seventh International RoboCup Symposium (RoboCup-2003)*, Lecture Notes in Artificial Intelligence. Springer, 2003.
- [Bonabeau *et al.*, 1999] Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, 1999.
- [Kawarabayashi *et al.*, 2005] T. Kawarabayashi, T. Kubo, I. Takahashi, J. Kuroiwa, T. Odaka, and H. Ogura. Adapting recognition of shootable situations by learning from experience and observation in a robocup simulated soccer game. In *Proceedings of the 3rd International Symposium on Autonomous Minirobots for Research and Edutainment (AMiRE 2005)*, pages 155–156. Springer, 2005.
- [Kazama *et al.*, 2005] Toshiya Kazama, Ken Sugawara, and Toshimori Watanabe. Traffic-like movement on a trail of interacting robots with virtual pheromone. In *Proceedings of the 3rd International Symposium on Autonomous Minirobots for Research and Edutainment (AMiRE 2005)*, pages 383–388. Springer, 2005.
- [Kitano and Asada, 2000] H. Kitano and M. Asada. The Robocup humanoid challenge as the millennium challenge for advanced robotics. *Advanced Robotics*, 13(8):723–736, 2000.
- [Kühne and Walter F. Lages, 2005] Felipe Kühne and João Manoel Gomes da Silva Jr. Walter F. Lages. Point stabilization of mobile robots with nonlinear

- model predictive control. In *IEEE International Conference on Mechatronics and Automation 2005*, 2005.
- [Langton, 1997] Christopher G. Langton, editor. *Artificial Life: An Overview*. MIT Press, 1997.
- [Nii *et al.*, 2005] M. Nii, M. Kajihara, Y. Takahashi, and T. Nakashima. An action rule discovery technique from simulated robocup soccer logs. In *Proceedings of the 3rd International Symposium on Autonomous Minirobots for Research and Edutainment (AMiRE 2005)*. Springer, 2005.
- [Nishino, 2005] Junji Nishino. Oz-rp project webpage. <http://www.fs.se.uec.ac.jp/nishino/ozrp/logs/>, 2005.
- [Riley and Veloso, 2001] P. Riley and M. Veloso. Coaching a simulated soccer team by opponent model recognition. In *Proceedings of the Fifth International Conference on Autonomous Agents*, 2001.
- [Sma, 2006] Robocup f180 league 2006 rules. <http://www.cs.cmu.edu/~brettb/robocup/rules/>, April 2006.
- [Yoshikawa, 2002] Kazuhiko Yoshikawa. Eco-Be!: A robot which materializes a watch's life (japanese). *Nature Interface*, (8):56–57, October 2002.

位置に基づく行動規則を実装した入門用サッカーエージェント作成キット OZED

The OZED introductory soccer simulation kit using KONOHEN-fuzzy rule-based reasoning agent system

西野順二、久保長徳、下羅弘樹、中島智晴

Junji NISHINO, Takenori KUBO, Hiroki SHIMORA, Tomoharu NAKASHIMA

電気通信大学、仁愛大学、福井大学、大阪府立大学

Dept. of systems engineering,

The University of Electro-Communications

nishino@se.uec.ac.jp

Abstract

This paper presents a RoboCup soccer simulation introductory development kit OZED. The OZED system contains a novel and user friendly agent decision rule system; Konohen-fuzzy reasoning system which is denoted by intuitive regional propositions. We discuss about the OZED system and the behavior representation ability of Konohen-fuzzy reasoning.

1 はじめに

RoboCup サッカーシミュレーションプレイヤー作成入門キット OZED(Open Zeng for EDucations) [西野 06]は、計算機やプログラミング知識の浅い中高生でもサッカーエージェントの構築と実験を可能にすることを旨としたシステムである。エージェントの作成作業を通じて、マルチエージェント環境におけるエージェントプログラミングの基礎についての教育的効果を与えることが目的であり、RoboCup サッカーシミュレーションリーグに適合したサッカーエージェントを、楽しく作成できるようにすることを目標としている。

シミュレーションリーグは、さまざまな知的手法を実践的に比較する研究用のプラットフォームであると同時に、演習などでロボットやエージェントを作成する標準教材としても使われている。これらの教材では、使用する技術の高度さと作業量から、これまでは上は大学院生下は工学部学生以上が対象であった。入門システムとして、行動ベースエージェントモデルに基づいた開発システム[Pau99]もあるが、設定すべき数値的パラメータ数も多く少なくとも学部学生程度のエージェントに関する知識が必要である。

本システムでは作業対象としてのRoboCupをそのままに、中高生や趣味でエージェントプログラミングを行い

たい非専門家を対象にした教育的な開発システムを目標としている。

とくに直観的な位置に基づくファジィルールに基づいた行動規則を、より直接的なビジュアル表現と組合せて編集可能なものとし、文字による操作をすることなくエージェントの行動を作成できるようにした。

2 OZED システム

OZED キットは、統合ランチャ、サーバ、バイナリプレイヤー、エージェントエディタ、エージェント実行サブシステム、サンプルルールズ、ruby/soccer ライブラリをまとめた、サッカーシミュレーションの入門キットである。手軽な行動設計や、実験を支援し、ruby によるライブラリを用いた高度な利用も可能としている。

2.1 OZ 的チームシステムによる協調実験環境

ヘテロジーニアスなマルチエージェント実験システムを作成するには、当然ながら多数のエージェントを準備しなければならないことが、開発上の負担となる。OZED システムではRoboCup ヘテロ協調チーム OZ の概念を継承して、この負担を軽減する。

チーム OZ[Nis01, 西野 00]は、サッカーシミュレーションリーグにおいてヘテロジーニアスなエージェントの協調を実験するために行ったプロジェクトである。複数の作者による異なったアーキテクチャのサッカーエージェントを「寄せ集め」てチームを構成しそれらの協調の様相を調べるものである。

サッカーは11体のエージェントで一つのチームとなる。OZED では、チーム構成に必要な他のプレイヤーエージェントとして、すべて一人のユーザが準備するのではなく、OZ のように他のユーザやシステム供給者が作成したプレイヤーを用いる。これによって、11体すべてを作る必要がなくなり、ユーザは1つのエージェントを集中的に開発、改良すれば良い。サッカーはポジションによる役割分

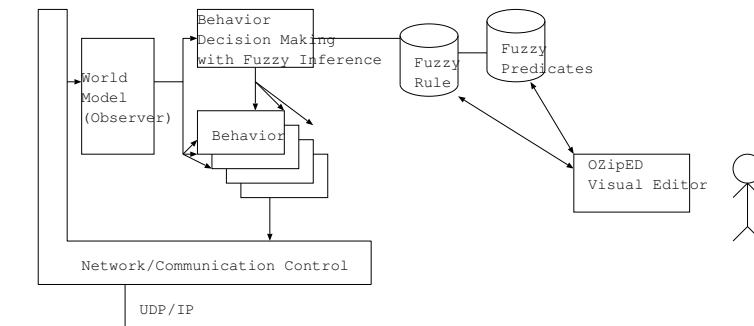


Figure 1: OZED システム (実行部分)

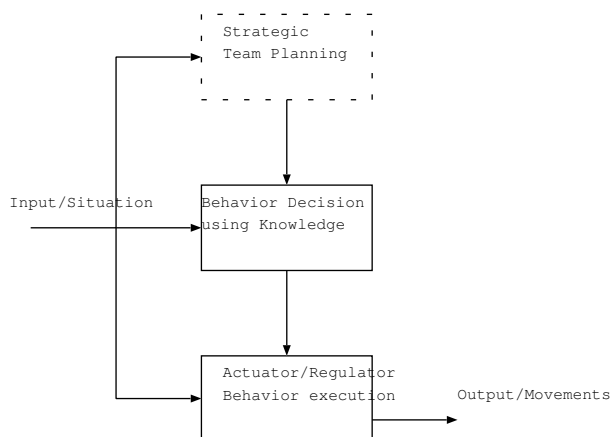


Figure 2: OZED エージェントモデル

担がよく行われ、OZED でも同様にユーザが各個人の好みのポジションを選び、そこでの行動を実装したエージェントを選択的に作ることが可能である。

2.2 実行部分の全体構成

OZED システムの実行部分の全体構成を図 1 に示す。大きく、行動エージェントとルールエディタの二つの部分から構成される。エージェントは実装上、ルールエンジン、ルール、ベースの三つからなり、それぞれ、このへんファジィ推論システム、サンプルルール群、ruby/soccer である。ルールエディタは、行動ルールとそこで使われるファジィ命題をビジュアルに編集できる機能を持ったユーザインタフェースである。

2.3 サッカーエージェントの意志決定モデル

本システムで提供するサッカーエージェントは、図 2 に示した典型的な階層型の意志決定モデルを採用している。

開発キットとして、状況認識機構と、制御階層の最下層にあたる各種行動をライブラリとして提供する。本システムの実装では構築が繁雑となる、再上層のチーム全体とも関わる戦略決定部は扱わないものとした。

中層部の行動決定部をファジィ推論によって実装し、ユーザはこの行動規則を構築・編集して意図するエージェント

行動を実現する。エージェント行動を作成する上で、ルールベースでの設計は負担が少ない。

このプレイヤー本体は ruby で記述されており、実装上は、OZED プレイヤルーチンは以下のループで動作する。

1. サッカーサーバと通信 (ruby/soccer)
2. 状態認識結果のベクトル化 (ruby/soccer)
3. 認識結果から行動選択 (ファジィ推論)
4. 選択行動の行動 (ruby/soccer)
5. 1 へ戻る

行動を決定する意志決定部分のファジィルール以外は、一般的なサッカーエージェントの様式に従っている。

2.4 ビジュアルルールエディタ

ビジュアルルールエディタは、OZip システムに端を発するこのへんファジィルール編集インタフェースである。OZip システムは、OZ プロジェクトの一つである OZ-RP システム[西野 02]の実装として作成された、人間パイロットがリアルタイム環境下で RoboCup サッカーシミュレーションサーバに接続して行動するためのインタフェースである。本エディタは、OZip と同様の画面構成をベースに、その上で「このへん」という位置を直接表現できるようにしている。ユーザインタフェースであるビジュアルエディタの動作のスナップショットを図 3 に示す。

2.5 ruby/soccer ライブラリ

バーチャルサッカーエージェントに必要な、状態認識、基礎動作やネットワーク通信機能などは、ruby を用いて作成され、エージェントライブラリを構成している。

ruby/soccer ライブラリが提供する機能は以下の通りである。

- ネットワーク通信機能：RoboCup サッカーサーバとの接続および通信を行う。サッカーサーバは UDP/IP 上のテキストベース独自プロトコルを持つため、本通信機能部では、通信の同期、バッファ管理、プロトコルパースを行う。



Figure 3: ビジュアルルールエディタ

- 状態認識：各エージェントが持つ状態オブザーバである。バーチャルフィールド上では、視野角、認識可能距離、サンプリング間隔などの制限がエージェントに課されている。また、各バーチャルサッカーエージェントがフィールド上で検知する環境情報は、サッカーサーバから独自プロトコルで送付される。本認識部では、プロトコルのデコードを行い、また時間的かつ空間的に補完・予測をする。本システムでは、シミュレーションステップ (0.1 秒) 単位で、各オブジェクトの位置を合わせた 46 次元ベクトルで出力を行う。
- 基礎動作ライブラリ：ドリブルやボールキープなどの単一目的で複合運動となる動作を実装したライブラリである。サッカーサーバが受け付けるエージェントの運動は、ダッシュ、キック、ターンと言った物理的にプリミティブなものだけである。ある地点へのドリブルはボールキックと移動の運動を交互に繰り返した複合運動であり、認識、スタミナ管理及びタイミング制御も含めた高次の実装が必要であり、行動マクロとして基礎動作ライブラリで提供する。

OZED で提供する基礎動作は次の通りである。

- gotoball: ボールを発見しその方向へ走る。
- shoot: ゴールの方向へキックを行う。
- centering: ゴール前のチームメイトへパスを行う。
- clearball: ボールに近付き直近の空いたサイドライン側へボールをキックする。

- holdball: 直近の相手プレイヤーと反対の側へボールを小さくキックし移動しつづける。
- positioning: 設定されたゾーンに戻る。
- go(x): フィールド上の点 x へ走る。
- kick(x): フィールド上の点 x へキックする。

2.6 サンプルルール群

プログラミングやシステムの構築技能の習得では、良質で豊富なサンプルに触れることが教育的効果を高める。OZED ではサンプルルールを組み入れ、できるだけ段階を追ってルール構築の仕組みを理解できるようにしている。

2.7 パッケージ

ユーザが構築したサッカーエージェントを試用するためには、自身のエージェント以外に、周辺環境を整える必要がある。最小構成では、サッカーサーバ、チームメイト、対戦相手、モニターが必要となる。サッカーサーバなどは教育研究目的でソース形式で公開されている。研究開発環境として RoboCup サッカーシミュレーションを用いる大学生・大学院生であれば、これらをソースから構築することも可能であるがその負担は多い。本システムが対象とする入門的なレベルの中高生では解決が困難である。そこでこれらを簡単に使用できるように単一のパッケージに組み入れた。現在は、UNIX (FreeBSD, Gentoo Linux), MacOS 10.4 で稼働する。OZED システムおよびエージェントは ruby/tk のスクリプトと、gtk+ または qt を用いたグラフィック環境で動作し、想定するプラットフォームでの利用が可能である。

2.8 サポートウェブ

サッカーエージェントを構築する動機付けを強めるため、OZEDのウェブを公開し、そこで相互にプレイヤーの交換や情報交換を促進する。

ウェブからサービスされる内容のうち、本システム、マニュアル、チュートリアル、サンプルの配布については、ウェブからの一方向性のサービスである。いっぽうユーザーエージェントの収集と再配布は、OZEDのユーザが作成した新規のエージェントの行動規則を集め再配布し、他のユーザが自由に使用できるようにするものである。これによって、互いに離れた環境でも相互にプレイヤーの交換を行い、試合を行うことができる。こうしたユーザ自身によるサンプル交換によって、より良い行動、より強い行動を行うエージェントを構築の手がかりをユーザ自身が学ぶことができるようにしている。

3 このへんファジィ

このへんファジィは、フィールド位置に基づいた直観的な行動規則を表現するシステムである。2次元の実空間を、このへんで、という位置に関する状態分割を行い、ファジィ推論エンジンを介して解釈・実行を行う。

3.1 このへんファジィ推論システム

位置に基づいたあいまいな概念を表現する命題と推論システムとして、2次元上のファジィ集合を用いたファジィ推論を用いる。

ファジィ推論による意志決定は、並列に列記されたルール群からなる推論方式であり、ルールの条件部には状態を変数とするファジィ命題を用いる。本システムでは「右サイド」「ゴール前」「自分の近く」など、サッカーフィールド上の物理的な位置の曖昧な領域をファジィ集合で表現し、「ゴール前ならシュートする」と言った直観的な行動規則を扱えるようにしている。

2次元上のファジィ集合を用いることにより、ある対象物Aの位置が $p=(x,y)$ の座標で与えられるとき、位置を表す2次元上のファジィ集合Gに対して、 $\mu_G(p)$ と単一の演算で表すことができる。従来のファジィ推論で用いられてきた手法では、条件部のファジィ集合として1次元上の集合の直積を用いていた。このため2次元上で見ると長方形領域のみしか表現できず、くの字形など凹凸のある領域は複数の長方形ファジィ領域の和として表現せざるおえなかった。本提案手法では「このあたり」という曖昧な2次元領域をそのまま2次元上の単一のファジィ集合により図上で指定できる。このため、より直観にあった位置条件命題の設定を可能としている。例として「右コーナー」を表す曖昧な領域を図4に示す。

また、これらのあいまいな領域はビジュアルエディタを用いてユーザによって直観的に定義される。エディタ上で

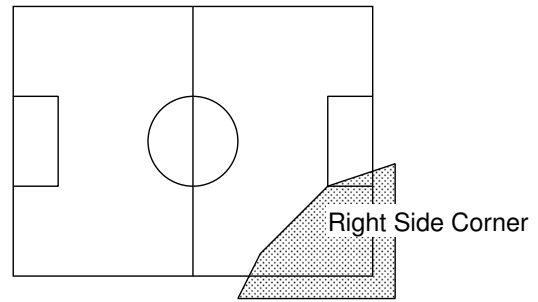


Figure 4: 右コーナーを表すファジィ領域

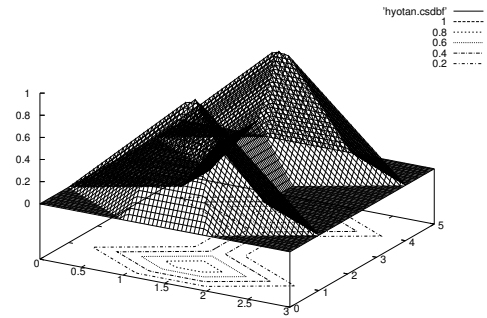


Figure 5: 複雑なアウトラインを持ったファジィ領域

領域のアウトラインを指定すると、そのアウトラインをもとにファジィ集合化される。凹凸部のある複雑なアウトラインを持ったファジィ集合の例を図5に示す。

3.2 推論原理

このへんファジィ推論は、ファジィ命題をもちいた並列発火型のルールベース推論である。ルールは以下の形式である。

$$R_i : \text{if } obj_i \text{ is } \tilde{A}_i \text{ then Do } Act_i \quad (1)$$

条件部の $obj_i \text{ is } \tilde{A}_i$ は2次元空間上のファジィ命題であり、具体的にはサッカーフィールド上の「このへん」という、主観的かつ直観的に定められたあるあいまいな領域 \tilde{A}_i に、 obj が適合するかについてのファジィ真値 $[0, 1]$ の値である。

対象 obj は、状態変数 $(X_1 X_2 \dots X_{22} B)$ のうちの一部であり、たとえば ball ならば $B = (B_x B_y) \in R^2$ の2変数を指す。あいまいな領域 \tilde{A} への適合度 g は、2次元上のファジィ集合 \tilde{A} のメンバーシップ関数 $\mu_{\tilde{A}}$ によって求まる。

$$g = \mu_{\tilde{A}}(X), X \in R^2 \quad (2)$$

推論結果の行動 Act_i は ruby/soccer で定義された基礎動作へのシンボルである。先に述べたエージェントモデル中で ruby/soccer がのライブラリが起動し、実際の行動が実行される。

3.3 推論結果の統合

このへんファジィ推論は、並列発火による推論を行う。ファジィ推論では、与えられた状態に対して、複数の規則がそれぞれ $g_i \in [1, 0]$ の範囲の重みを持って適合する。結果の統合は各推論規則 R_i の適合度 g_i をもとに行う。現在の OZED では、最大適合度を与える規則の行動を採択する。

$$Result = \{Act_j | g_j = \max_i(g_i)\} \quad (3)$$

適合度が同じ場合の競合解消は、ルール番号の小さいものをとることとした。

この他の統合手法としては、適合度に応じた確率的に 1 つ選択、閾値以上のうちから 1 つ選択、適合度に応じた混合などがある。これらの実行方法についてはファジィアルゴリズム[Zad68, Tan75]において定式化されている。

3.4 このへんファジィ規則の表現能力

このへんファジィによる位置に基づくファジィ推論は、記憶を持たないエージェントの任意行動規則を表現可能である。

記憶を持たないエージェントの行動規則は、すなわち、現在の状態 s に対して行動 Act を割り当てる任意の写像 $\pi : S \mapsto A$ である。

いっぽう、このへんファジィ集合を適当な大きさの範囲で多数作成し、それぞれのメンバーシップ関数が互いに重なる境界を調整することで、任意のルール適合度 $0 < e < 1$ を閾値として、状態空間 S 全体を細分化した状態 \tilde{s}_i の有限な集合和で表現することができる。

$$S = \bigcup (\tilde{s}_i) \quad (4)$$

ここですべての細分化した状態 \tilde{s}_i ごとに行動 Act_i を割り付ける規則

$$R_i : \text{if } s \text{ is } \tilde{s}_i \text{ then Do } Act_i \quad (5)$$

を作れば任意の行動規則を表現する写像となる。

3.5 相対的表現について

OZED では条件部の表現として相対位置も許している。相対位置を用いた表現を含めてもあらゆる行動規則を表現できる。これは絶対位置による規則で表現できるものを、その簡略化したものととらえることができる。

たとえば「ボール (ball) が自分 (self) の前 (B) にある」というファジィ命題中の「自分の前 (B) にある」という相対的な表現は、

$$\text{if ball is selfRelativeB then} \dots \quad (6)$$

と書ける。ここで、これは自分の位置 (self = A = (X_x^s, X_y^s)) から決まる「前」の位置 $B(A)$ をそれぞれ取る一連の展開されたルール群で表現できる。

$$R_i^{resolved} : \text{if self is } A_i \text{ and ball is } B(A_i) \text{ then} \dots \quad (7)$$

逆に、この展開されたルール群を縮約した表現が相対表現である。

3.6 ルール数の縮約可能性

細分化した状態 $\bigcup (s_i) = S$ にもとづいて、作られた完全表現規則群があるとし、このとき後件部の行動 $Act_j \in A$ は有限で m 個とする。ここで、規則群を行動の添字ごとに集めた規則集合 $R_j = R_i | Act_i = Act_j$ を行動の数だけ作ることができる。すると各規則の条件部のファジィ命題のファジィ集合を統合できる。

$$\tilde{A}sum_j = \bigcup_{R_i \in R_j} \tilde{A}_i \quad (8)$$

ここで演算 \cup はファジィでの集合和である。

完全表現した規則群を行動ごとに統合することで、 m 個のルールで表現できる。

$$R_i : \text{if state is } \tilde{A}sum_i \text{ then do } Act_i \quad (1 \leq i \leq m) \quad (9)$$

4 システムの評価と考察

OZED で提案する位置に基づいたエージェント行動規則の記述能力と、ruby による実装と速度について検討を行う。

4.1 行動規則の記述力

このへんファジィによる行動規則表現は原理的に、任意の記憶を持たないエージェントの行動規則を表現できることを示し、ファジィ規則の並列発火および実行の方式に起因する統合可能という性質から、行動によって規則を縮約することができることを示した。

しかしながら実用上は、リアルタイム制約からあまり多数の規則の実行は不可能である。また、理解不能な縮約を行った規則群では、直観的に分かりやすいことが主たる利点であるファジィ推論としての意義が薄くなってしまふ。

実際の入門キットとして使用するためには、10 から 20 程度のあまり多くない規則数で、十分な行動をするエージェントが表現できることが望ましい。

OZED キットのサンプルとして掲げたエージェント 7 種類は、それぞれ 1, 6, 5, 2, 7, 4, 4 本の規則で構成されている。最大 7 規則、平均 4 規則で、FW や MF としてのポジショニングや行動を行っている。これらのプレイヤーの行動が十分と言えるかどうかの判断はチームとしてのバランスとも関わるものである。このため十分性を断言することはできないが、プレイを観察する範囲では最低限で必要な判断と行動を行っていた。

4.2 実行速度

エージェント全体がrubyで作られているため、実行速度は速くはなく、場面によっては推論速度が問題となる。サッカーサーバのシミュレーションステップは100ミリ秒単位である。このため、自作サッカーエージェントを用いて試合を行うとき、一回の行動決定ステップが通信を含めて100ミリ秒未満に収まるかどうかの評価の閾値である。また、サッカーサーバやモニタ、チームメイト、相手チームエージェントを同時に稼働させると最大で23プロセスとなるため、1エージェントあたり4.3ミリ秒が目安となる。

rubyによるファジィ推論全体での1ステップ分の演算時間は、201.1ミリ秒であり、このうちファジィ集合に関する部分の計算量を除いた、ルールの処理や通信処理などは0.88ミリ秒であった。その演算時間のほとんどをファジィ集合処理が消費していた。

この速度上の問題を回避するため、家庭などで用いられる配布用OZEDキットでは、5対5のチーム構成とし、プロセス数を減らすこととした。また、負担の重いファジィ集合処理については、C言語によって実装されたrubyの拡張ライブラリとすることで高速化を計った。

これらの対策により、MacOSX(CPU G4 1.45GHz)において、1ルール処理が201.1ミリ秒から1.0ミリ秒に改善され、実用上の問題は解決した。

ビジュアルルールエディタはC++/Gtk+で開発され十分な応答である。また、エディタ作業画面は、エージェント実行ほど時間制約が厳しくなく、多少の応答遅れがあっても使用上の不都合はほとんどない。

5 まとめと今後の課題

このへんファジィは人間が通常持っている、位置という直観的な概念をそのまま表現できる仕組みであり、その利用も簡単である。本稿では原理的な表現能力について示した。キットとして実装したOZEDでは、実用的な範囲として規則数がある程度しぼってもユーザの意図する動きを実現でき、実行速度も十分であることを確かめた。

提案システムで実装した位置に基づいたファジィ命題の使用は、ユーザが直観的に行動の意味を理解し開発を進める助けとなる。実験対戦の相手チームや、自チームのチームメイトなどをパッケージとして組み込んだことにより、サッカーサーバを起動するための統合環境とあいまって、経験や知識の少ない中学生、高校生でも公式のRoboCupに自宅で触れることを可能にした。さらに、内部のエージェントシステム本体はRubyベースで作られているため、これらをライブラリとして用いることで、オブジェクト指向エージェント開発にまで立ち入ることができる。

今後は、このへんファジィ推論の表現能力について実用的な範囲での規則数における精密な議論が必要である。

なお、本システムの開発は、独立行政法人情報処理推進機構(IPA)より、2005年度上期未踏ソフトウェア創造事業の支援により行われたものである。

参考文献

- [Nis01] Nishino, J., T. Morishita, and T. K. and: *RoboCup 2000: Robot Soccer World Cup IV*, chapter Open Zeng: An Open style distributed semi-cooperative Team development project from Japan, pp. 473–476, Springer, 2001.
- [Pau99] Paul Scerri, A. T., S. Coradeschi: *RoboCup-98: Robot Soccer World Cup II*, chapter A User Oriented System for Developing Behavior Based Agents, pp. 173–186, Springer, 1999.
- [Tan75] Tanaka, K. and M. Mizumoto: Fuzzy Programs and Their Execution, in etc., L. A. Z. ed., *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, pp. 41–76, New York US., 1975, Academic Press.
- [Zad68] Zadeh, L. A.: Fuzzy Algorithms, *Information and Control*, Vol. 12, pp. 94–102, 1968.
- [西野 00] 西野, 森下, 木下, 鈴木他: シミュレーションドリームチーム OZ の挑戦, AI チャレンジ研究会 第 9 回資料, pp. 16–19 人工知能学会, 2000.
- [西野 02] 西野: ドリームチーム OZ と人間チーム OZ-RP の挑戦, 日本ロボット学会誌, Vol. 20, No. 1, pp. 39 – 40, 2002.
- [西野 06] 西野, 久保, 下羅, 中島: ロボカップサッカーシミュレーションエージェント開発体験キット OZED, 第 4 7 回プログラミングシンポジウム報告集, pp. 79–86 プログラミングシンポジウム委員会, 情報処理学会, 2006.

ニューラルネットワークによるゲームログからの行動獲得

Behavior Acquisition by Neural Networks through Game Logs

並川直樹, 小野玄基, 横山智史, 高谷将裕, 中島智晴, 石渕久生

Naoki NAMIKAWA, Genki ONO, Satoshi YOKOYAMA,

Masahiro TAKATANI, Tomoharu NAKASHIMA, Hisao ISHIBUCHI

大阪府立大学

Osaka Prefecture University

{namikawa@ci., ono@ci., sts86y@ci., takatani@ci., nakashi@, hisaoi@}cs.osakafu-u.ac.jp

Abstract

This paper proposes a framework for acquiring a low-level behavior of a soccer agent. The task of a learning agent is to mimic the behavior of a target agent with a well-trained behavior. Neural networks are used to represent the behavior of the target agent. In order to obtain a set of training data, we convert game logs of the target agent into a set of input-output pairs for the neural networks. We show the effectiveness of the proposed framework through the computational experiments.

1 はじめに

RoboCup は, ロボット工学や人工知能など, 多岐に渡る技術の融合・発展のための標準問題の提供を目的として提唱された国際的プロジェクトである. 協調マルチエージェント問題を主に取り扱う RoboCup シミュレーションリーグでは, ソフトコンピューティングに基づいた手法がよく適用されている. 例えば, 中島らは, ファジィQ 学習によりオンラインでボールをインターセプトするスキルを獲得する手法 [中島, 2003] や, 進化型計算によりチーム戦略を獲得する手法 [Nakashima, 2004] を提案している. また, Luke ら [Luke, 1996] は, 共進化の概念を取り入れた遺伝的プログラミングを提案しており, 進化の初期段階では団子サッカーであったチーム戦略が, 進化の終了段階では, フォーメーションサッカーを行うような戦略へと進化することを示している.

通常, 優れたドリブルをプレーヤに実行させるためには, 開発者自らの経験やノウハウを生かして, ハンド・コーディングを行わなければならない. 本論文では, ニューラルネットワーク (以下 NN) によりドリブルスキルを獲得する手法を提案する. 提案手法では, 優れたドリブルを行

うプレーヤのログデータを利用して NN を学習させることにより, ドリブルスキルが獲得される.

2 ニューロ・ドリブル

本論文では, プレーヤが NN を用いてドリブルを行う動作をニューロ・ドリブルと呼ぶことにする. ニューロ・ドリブルを実装するため, 優れたドリブルを行うチームのログデータから学習用データを抽出し, バックプロパゲーション法 [Rumelhart, 1986] により NN の学習を行う. プレーヤは, 学習により得られた NN を用いてニューロ・ドリブルを行う.

本論文では, 2 種類のニューロ・ドリブルを実装するための方法を提案する. 一つは, ドリブルを構成する三つの基本動作 (キック, ダッシュ, ターン) に対して, それぞれ独立した NN を構築し, 各行動のパラメータを NN により決定してドリブルを行う方法である. もう一方は, 単一の NN を用いて基本行動の選択, 及び選択した行動のパラメータの値を NN により決定してドリブルを行う方法である. 以下では, 学習用データの生成手順, 本論文で用いた NN の構造, 及びニューロ・ドリブルを実装するための手法を説明する.

2.1 学習用データの生成

優れたニューロ・ドリブルを実現するためには, 学習用データとして優れたドリブルを行うプレーヤのデータが必要となる. 本論文では, 2004 年の RoboCup 世界大会で優勝したチームである STEP のドリブルを NN に学習させる. STEP は優れたドリブルを特徴としている. 学習用データを生成するために, STEP と他のチームを対戦させ, 試合のログデータを収集する. 試合のログには, プレーヤがドリブルを行っている際のフィールド上の情報とプレーヤがサーバに送信した情報が含まれているため, NN の学習用データとなる知覚・行動データを抽出することができる. 以下に学習用データの生成手順を示す.

[学習用データの生成手順]

- Step 1: 学習対象のプレーヤを含むチームによる試合を複数回行い、試合のログを収集する .
- Step 2: 収集したログから、学習対象のプレーヤがドリブルを行っている部分を抽出する .
- Step 3: ドリブル部分を抽出したログから、サイクルごとに環境の状態に関する知覚情報を入力値、そのサイクルでプレーヤがとった行動を出力値として学習用データを生成する .

2.2 ニューラルネットワーク

本論文では、Figure 1 に示すような 3 層階層型ニューラルネットワークを用いる . n 次元ベクトル $x_p = (x_{p1}, x_{p2}, \dots, x_{pn})$, $p = 1, 2, \dots, N$ が NN への入力として与えられたとすると、各ユニットの入出力関係は以下のようになる .

$$\text{入力層} : o_{pi} = x_{pi}, \quad i = 1, 2, \dots, n_1 \quad (1)$$

$$\text{中間層} : o_{pj} = f(\text{net}_{pj}), \quad j = 1, 2, \dots, n_2 \quad (2)$$

$$\text{net}_{pj} = \sum_{i=1}^{n_1} w_{ji} o_{pi} + \theta_j \quad (3)$$

$$\text{出力層} : o_{pk} = f(\text{net}_{pk}), \quad k = 1, 2, \dots, n_3 \quad (4)$$

$$\text{net}_{pk} = \sum_{j=1}^{n_2} w_{kj} o_{pj} + \theta_k \quad (5)$$

ここで、 N は学習用データの数、 n_1, n_2, n_3 はそれぞれ、入力層、中間層、出力層のユニット数、 w_{ji}, w_{kj} はユニット間の結合強度、 θ_j, θ_k はしきい値を表している . 但し、入出力関数 $f(x)$ は次のようなシグモイド関数であるとする .

$$f(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

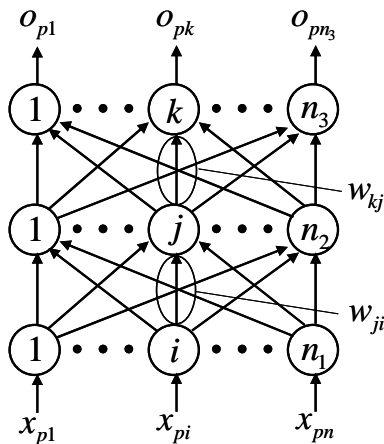


Figure 1: 3 層階層型ニューラルネットワーク

また、NN の学習が進んでいることを確認する指標の一つとして、自乗誤差がある . 自乗誤差は、学習用データに対する NN の出力値の誤差であり、次のように表される .

$$E = \sum_{p=1}^N E_p = \sum_{p=1}^N \sum_{k=1}^{n_3} \frac{(t_{pk} - o_{pk})^2}{2} \quad (7)$$

ここで、 t_{pk} は学習用データを表している . NN の学習は、自乗誤差が最小となるように、バックプロパゲーション法により行われる .

2.3 複数の NN によるニューロ・ドリブル (手法 1)

Figure 2 に、複数の NN によるニューロ・ドリブルで用いる場合の構成を示す . 手法 1 では、三つの基本行動 (キック, ダッシュ, ターン) に対してそれぞれ一つの NN を用意する . 各 NN は、プレーヤの視覚情報を入力値として行動に関するパラメータを出力する . キック NN はボールに加える力と方向の 2 出力、ターン NN はターンの方向の 1 出力、ダッシュ NN はダッシュの大きさの 1 出力を出力する .

プレーヤは、ボールが蹴ることのできる範囲に入ったときに、ニューロ・ドリブルを開始する . 手法 1 において、基本行動の選択は予め設定した条件に基づいて行われ、選択された行動に対応する NN が出力するパラメータの値に従ってプレーヤは行動する . 以下に、手法 1 におけるニューロ・ドリブルの手順を示す .

[複数の NN によるニューロ・ドリブルの手順]

- Step 1: ボールを足元で止める .
- Step 2: ターン NN を用いてドリブルを行う方向へ体を向ける .
- Step 3: ボールがプレーヤの中心より前にある場合は、ダッシュ NN を用いてダッシュを行う . ボールがプレーヤの中心より後にある場合は、キック NN を用いてキックを行う .
- Step 4: 終了条件が満たされていれば、手順を終了する . それ以外の場合は、Step 3 へ戻る .

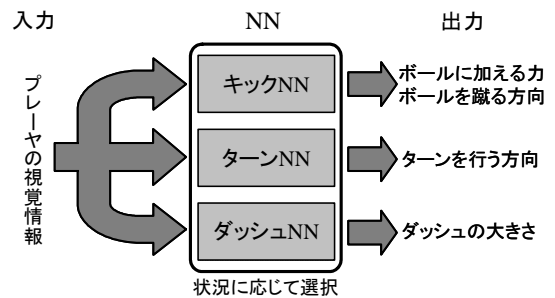


Figure 2: 手法 1 の NN の構成

Table 1: 実験設定

	入力ベクトル	ユニット数
手法 1	プレイヤーの絶対速度と体の向き ボールのプレイヤーに対する相対座標と相対速度 近隣の敵 3 人のプレイヤーに対する相対座標	入力層：13 中間層：30 出力層：各 NN の出力数
手法 2	プレイヤーの絶対座標と絶対速度と体の向き ボールのプレイヤーに対する相対座標と相対速度 近隣の味方 1 人と敵 2 人の相対座標，相対速度，体の向き	入力層：24 中間層：60 出力層：7

2.4 単一の NN によるニューロ・ドリブル (手法 2)

Figure 3 に，単一 NN によるニューロ・ドリブルで用いる NN の構成を示す．手法 1 とは異なり，手法 2 では，基本行動の選択及び各行動に関するパラメータの値といったすべての意思決定が，単一の NN によって包括的に行われる．行動を選択するために，キック，ダッシュ，ターンの各基本行動に対する出力が存在する．プレイヤーは，これらの値を比較し最も大きな値をとる行動を選択し，対応するパラメータの値に従って行動する．手法 1 と同様に，プレイヤーはボールが蹴ることのできる範囲に入った場合に，ニューロ・ドリブルを開始し，終了条件が満たされるまでこれを継続する．

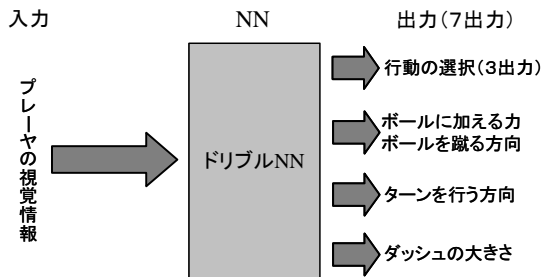


Figure 3: 手法 2 の NN の構成

3 数値実験

3.1 実験設定

提案手法の有効性を検証するため，数値実験により各手法の性能を調べた．NN の学習では，学習係数を 0.1 としてバックプロパゲーション法による重みの学習を行った．各手法に対する入力ベクトル，及び各層のユニット数を Table 1 に示す．また，問題の簡略化のため，プレイヤーがフィールド上の全情報を取得できるようにし，さらに，送受信する情報にノイズが含まれないように設定した．ニューロ・ドリブルの終了条件は，敵が近くに存在する場合，ボールが味方もしくは敵にとられた場合，スタミナが基準値以下となった場合，ボールが規定の場所に到達した場合，15 サイクル以上ボールから離れている場合とした．

3.2 複数の NN によるニューロ・ドリブルの性能

学習用データに対するキック NN の自乗誤差を Figure 4 に示す．学習の前半では，自乗誤差の値が大きく，プレイヤーはボールを蹴らずに，その場で留まっているだけであったが，学習の後半ではドリブルの動作を観測することができた．また，Figure 5 に手法 1 を用いたニューロ・ドリブルの様子を示す．Figure 5 より，手法 1 を用いるとドリブルスキルが獲得されることがわかる．

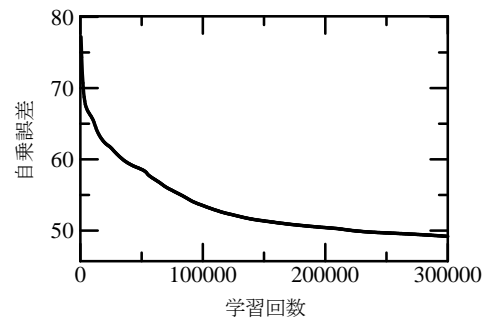


Figure 4: キック NN の自乗誤差

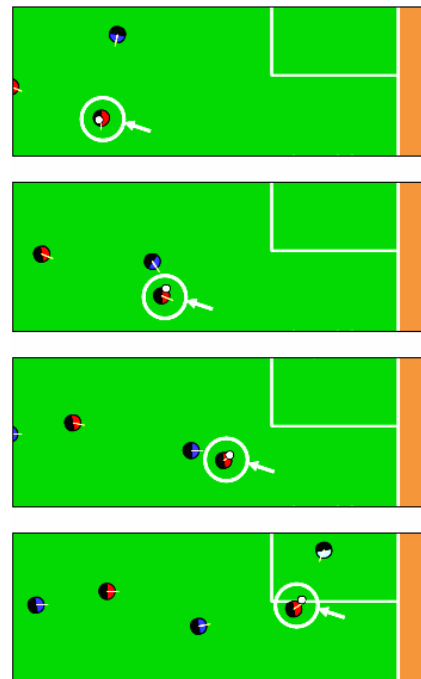


Figure 5: 手法 1 を用いたニューロ・ドリブル

3.3 単一の NN によるニューロ・ドリブルの性能

学習用データに対する NN の自乗誤差を Figure 6 に示す。手法 1 の場合と同様に、学習前半ではドリブルの動作は見られなかったが、学習が進んでいくと、ドリブルの動作を観測することができた。Figure 7 に、単一の NN によるニューロ・ドリブルと学習対象である STEP のプレイヤーによるドリブルを比較した図を示す。Figure 7 より、手法 2 を用いると学習対象の STEP と類似したドリブルスキルが獲得されることがわかる。

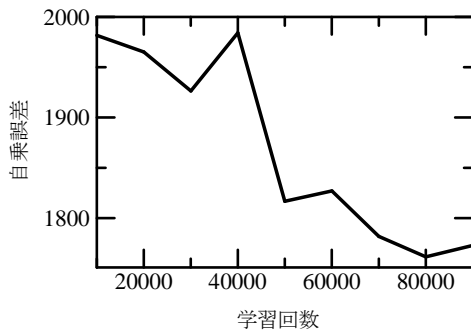


Figure 6: 単一の NN の自乗誤差

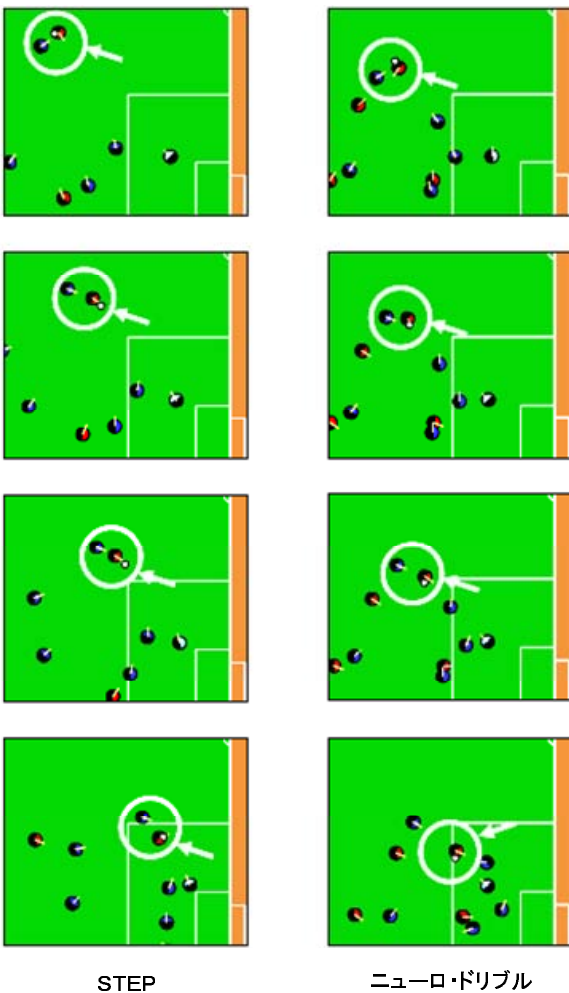


Figure 7: STEP とニューロ・ドリブルの比較

4 ニューロ・ドリブルの問題点

前章では、NN を用いてドリブルスキルが獲得されることを示したが、一方で攻撃方向とは逆へボールを強く蹴る動作も確認された。特に、行動の選択を NN により行っている手法 2 においてこのような動作が多く観察された。この動作の原因として、学習用データの不足が考えられる。例えば、STEP とニューロ・ドリブルを行うチームのポジショニングや戦略が異なるため、学習用データには存在しない状況がニューロ・ドリブル実行中に生じる場合がある。この問題の解決策として、さらに多くの学習用データを収集すること、ドリブルの行動として明らかに適切でない行動をとらないように条件を追加すること等が考えられる。また、NN の入力値や実装方法に関する検討も必要である。

5 おわりに

本論文では、NN を用いてドリブルスキルを獲得する手法を提案した。

NN の学習は、優れたドリブルを行うプレイヤーを模倣するように行われた。数値実験により、提案手法を用いると学習対象である STEP のプレイヤーと類似したドリブルスキルが獲得されることを示した。一方で、攻撃する方向とは全く逆へボールを強く蹴ってしまう動作も確認された。今後の課題は、このような問題を解決し、学習対象のプレイヤーのドリブルにより類似したニューロ・ドリブルを実現することである。また将来的には、ドリブル以外のパスやシュートといったスキルを NN により獲得する手法の開発を検討している。

参考文献

- [Rumelhart, 1986] Rumelhart D. E., McClelland J. L. and the PDP Research Group: *Parallel Distributed Processing(1)*, pp. 318-364, MIT Press, 1986.
- [Luke, 1996] S. Luke and L. Spector: Evolving Teamwork and Coordination with Genetic Programming, *Proc. of the First Annual Conference on Genetic Programming*, pp. 150-156, 1996.
- [中島, 2003] 中島智晴, 有働昌代, 石渕久生: ファジィQ 学習によるサッカーエージェントの行動獲得, *日本知能情報ファジィ学会誌*, Vol. 15, No. 6, pp. 702-707, 2003.
- [Nakashima, 2004] T. Nakashima, M. Takatani, M. Udo and H. Ishibuchi: An Evolutionary Approach for Strategies Learning in RoboCup Soccer, *Proc. of 2004 IEEE International Conference on Systems, Man and Cybernetics*, pp. 2023-2028, 2004.

ad-hoc ネットワークにおける音波を用いた端末間距離測定に関する一考察

Measurement of Distance Between Nodes in Ad-hoc Network by Acoustic Wave

植村 渉

Wataru UEMURA

龍谷大学理工学部

Ryukoku University

wataru@rins.ryukoku.ac.jp

Abstract

On ad-hoc network, distance information between nodes is needed for making routing tables. Usually we get information of a wireless signal power from a wireless network card because there is a relationship between the wireless signal power and distance between nodes. But some network cards do not provide us a wireless signal power or such information. In this paper, we provide a novel distance measurement method using acoustic wave. Since wave is not a monotone decreasing because it has 'nodes' which amplitude are always zero per wavelength. So it is too hard for us to get distance information from a wave signal power. In this paper, we propose the distance measurement method with the two or more acoustic waves. And experiments show its implementability.

1 はじめに

基地局を介さず、個々の端末同士が通信することでネットワークを構築するアドホックネットワークは、通信のためのインフラの設備や準備を必要としないことから、近年注目されている [1][2][3][4][5]。また、アドホックネットワークは、その性質上ネットワークが不安定であるため、ネットワークを安定化させる研究 [6][7] や、ネットワークを流れる情報であるパケットの経路設定に関する提案 [8][9][10][11] など、様々な研究が行われている。特に、お互いの端末（ノード）間の距離情報を知ることが、アドホックネットワークにおいて重要である [12][13][14]。ただし、ネットワークは随時変化するため、距離情報は必ずしも正確である必要はなく、大雑把な情報であっても、

十分有用である。ノード間の距離情報は、ネットワークに端末がどのようにつながっているかを知るために重要であり、通信のために用いる情報であるため、一般的には無線通信時の電波強度を用いて、お互いの端末間距離を知る [15][16][17]。無線通信を行う際、ハードウェアレベルにおいて、電波強度の情報は受信信号を扱う際に容易に獲得することができる。しかし、ソフトウェアレベルにおいては、下位のハードウェア寄りのドライバやチップに依存し、必ずしも獲得できるとは限らない。本研究では、アドホックネットワークの端末として、ソニーの四足歩行ロボットを使用する場合を想定している。このロボットに搭載している無線カードからは、電波強度情報を得ることができるが、それがどの端末との通信時に得られたかという情報は得られない。通常は、アクセスポイントとの通信になるため、これだけの情報で十分であるが、アドホックモードにおいては、端末間通信を行うため通信相手が複数になり、この情報だけでは、端末間距離が判定できない。また、アドホックモードでは、各端末がアクセスポイントと同等になるため、自分自身のポイントの情報として拾う場合もある。この場合も、通信相手との電波強度を知ることができず、端末間距離が得られない。

本研究では、四足歩行ロボットが、モノラルのスピーカと、ステレオのマイクを搭載していることに着目し、音波を用いて端末間の距離を測定する方法を提案する。そして、実験により、この手法の有効性を確認する。

以下、二章では、アドホックネットワークについて述べ、音波による端末間の距離の測定方法について提案する。三章では、実験に先立ち、実験に用いる端末である四足歩行ロボットの紹介と、それを用いたロボカップについて述べる。そして、四章にて実験を行い、五章でまとめる。

2 アドホックネットワーク

アドホックネットワークでは、従来の基地局型ネットワークとは異なり、中心となる基地局を用いず、各端末間で通信を行い、ネットワークを構成する。そのため、ネットワークインフラの構築が簡単であり、災害現場などにおいて大いに期待されている。ただし、その特性上、ネットワークは常に変化する可能性を持ち、従来のネットワークと比べると不安定である。

アドホックネットワークを構築する様々な方法があるが、最も簡単な構築方法のひとつとして、IEEE 802.11 のいわゆる無線 LAN ネットワークのアドホックモードを用いる方法がある。このモードを用いれば、中心となるアクセスポイントを設置することなく、通信可能圏内の他の同一設定のノードと簡単に通信することができる。IEEE 802.11 の中でも、最高 11Mbps の通信速度である 802.11b に対応した機器は、一般に広く出回っており、無線 LAN に対応しているノートパソコンでは、ほぼ全てにおいて使うことができる。

本研究で対象とする無線端末は、ソニー社製の四足歩行ロボット（アイボ¹）である。このロボットは、802.11b に対応した無線 LAN カードを搭載しており、アドホックモードを用いた通信をすることができる。また、ロボット自ら移動することができるため、移動型の通信端末としても期待できる²。

ネットワークのトポロジーが常に変化するアドホックネットワークでは、安定なネットワークである有線で用いる経路設定は、そのまま適用することは難しい。アドホックネットワークで用いられる経路設定の方法は、大きく分けて二種類ある。

ネットワークの状況を常に把握し、常時ルーティングテーブルを更新することで、経路設定を可能とする方法を、プロアクティブ型と呼ぶ。この方法では、常に状況を把握するために、トラフィックが増加する危険性がある。また、経路設定を行うときにだけ、ネットワークの状況を調べる方法を、リアクティブ型と呼ぶ。この方法では、通信要求が生じてから、ネットワークを調べるため、通信を行うまでの待ち時間が生じる。代表例として、前者の方式は、OLSR(Optimized Link State Routing)[8] や TBRPF(Topology dissemination Based on Reverse-Path Forwarding)[9] があり、後者の方式は、AODV(Ad hoc On-demand Distance Vector)[10] や DSR(Dynamic Source Routing)[11] が挙げられる。本研究においては、経路設定は既存の方式を用い、特に問題としないため、いずれの方式にも適用できる³。

¹ アイボはソニーの登録商標。

² 残念ながら、2006 年 3 月をもって生産中止となった。

³ ただし、人間の可聴域の音波を用いる以上、ネットワークの状況を常に調べるプロアクティブ型は、人間に対して適していないと考えられる。

2.1 端末間距離の推定

ルーティングテーブルを作成するためには、どの端末と通信できるか、また、その通信可能端末との距離はどれくらいを知る必要がある。通常は、通信時の受信電波強度を用いて、お互いの端末距離を知ることができる [15][16][17]。しかし、本研究で対象とする四足歩行ロボットでは、獲得した電波強度が、どの端末との通信における電波強度が知ることができない。また、アクセスポイントを用いた通信を行う場合においても、アクセスポイントとの電波強度は測定できても、他の端末とは直接通信しないため、端末間距離を知ることができない。そこで、本研究では、四足歩行ロボットがモノラルスピーカとステレオのマイクを搭載していることに着目し、音波を用いて端末間距離を測定する方法を検討する。実際に利用する際には、まわりからの雑音や、ロボット自身の雑音が影響を与える。また、壁などによる反射や吸収、障害物の回折など様々な要因も存在する。今回は、音波利用の可能性を検討するため、ロボットは動かさず、また環境もできるだけ静かな環境を選び、雑音などの要因は今後検討することとする。

距離を測定する方法として、超音波を用いる研究が多い [18][19]。音波を用いた位置の推定は、三次元空間で行うため厳密には三点用いる必要がある [20]。また、マイクに指向特性を持つコンデンサマイクを用い音源を定位する方法もある [21]。ただし、この方法は、四つのマイクを配置したタワーを二組用いるため、マイクの数が決まっているロボットへの適用は難しい。

本研究では、四足歩行ロボットに実装することを目的とするため、できる限り計算量を抑える必要がある。そこで、単純に、一方の端末が音を出し、もう一方の端末で音を受信し、その受信強度から距離が測定できないか検討する。

この時、解析が単純な音として、単一の周波数成分を持つ音を送信し、受信側ではフーリエ変換することで、周波数成分を抜き出し、フーリエ係数から音の強さを測定することを考える。ただし、単一の周波数成分を持つ波の場合、振幅は距離に対してサイン波を描くため、距離に対して単調減少にならない。直接波のみを検討する場合、振動しない「節」と呼ばれる点が生じる。そこで、本研究では、受信強度が距離に対して単調減少となるよう、以下の二つの方法を検討し、効果を確かめる。

まず、ステレオマイクを搭載していることから、音源とそれら二つのマイクとの距離に差が生じることを見越し、二つのマイクの受信強度の平均を用いる方法を提案する。次に、節となる場所は、周波数に依存するため、複数の周波数の音を用いることで、節の効果を除去し、距離を推定する方法を提案する。これらの効果を、実験にて確認する。

3 四足歩行ロボット

実験に入る前に、本節では、四足歩行ロボットの仕様について説明し、用いる対象環境を定義する。

本研究では、ソニーの四足歩行ロボットであるアイボを用いる。アイボには数種類の製品があるが、その中でも、ERS-7 と呼ばれるタイプを用いる。

ERS-7 には、IEEE802.11b に対応した無線 LAN カードが搭載されており、アクセスポイントを介して通信するインフラストラクチャモードと、直接端末同士で通信するアドホックモードに対応している。一般に IEEE802.11b は、11Mbps を使う場合、通信距離は見通し距離で 40m と言われている。また、壁約二枚を超えると通信が確立できなくなる。本研究では、広い空間での使用として、30m 程度の距離を仮定する。また、狭い空間での使用として、ロボカップ四足リーグ [23][24] のフィールドを仮定する。四足リーグのフィールドは、2006 年は 6m × 4m であるため、フィールド上の最長である対角線約 7m を最大距離と仮定する。

マイクは、頭の耳に当たる位置に設置されており、右耳、左耳に対応してステレオで音を拾うことができる。プログラミングは、OPEN-R[22] を用いた C++ 言語を使う。この場合、16ms 単位で、サンプリングレート 16kHz、サンプルサイズ 16 ビット、ステレオの PCM データ形式で音を獲得できる。ここでは、32ms 分データを蓄え、512 個のサンプリング点を解析に用いた。

スピーカは、お腹の部分に存在するが、基本的に足を使って立っている姿勢では、前方を向くようになっている。再生する PCM データは、サンプリングレートが 8kHz、サンプルサイズが 8 ビットのモノラルである。

移動に関しては、モータを直接制御して移動する方法や、モーションデータを用いて移動する方法がある。ここでは、本研究の本質ではないため、移動方法は限定しない。ただし、移動に伴う雑音が生じるため、できる限り雑音の生じない移動方法が望まれる。

4 実験

まず、再生する音の種類について予備実験を行い、再生する音の波形を決定する。次に、二つのマイクの受信強度の平均により、節を除去する方法を実験する。そして、複数の周波数を用いることで、節の影響を軽減する方法を実験する。実験方法は、二体のアイボを向かい合わせに置き、一体から音を再生し、もう一体で受信し、信号強度を測定し、十回平均を音圧として記録した。再生側の音の強さは、推奨される最大値である -10dB を用いた⁴。いずれのロボットも、図 1 の体勢で行う。この台車の先端を距離の

⁴ ハードウェアとしての最大値は、0dB である。しかし、0dB で音を鳴らし続けると出力アンプ IC の発熱でハードウェアが損傷する恐れがある [25]



Figure 1: a 4-legged robot.

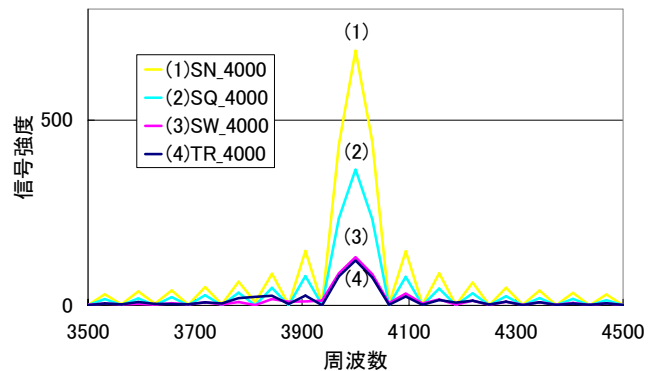


Figure 2: a frequency analyze at 440Hz (distance = 1m).

原点とし、先端と先端の距離を実測値とした。まず、1m の距離に二体のロボットを置き、周波数 4kHz において複数の波形を再生し、受信音波のスペクトルを調べた。サイン波 (SN4000)、ノコギリ波 (SW4000)、方形波 (SQ4000)、そして三角波 (TR4000) を用いた。結果は、図 2 である。対象となる 4kHz 前後を表示している。受信した信号の強度は、サイン波が一番高く、方形波が半分ほどであり、ノコギリ波と三角波がほぼ同じ結果となった。同様の実験を 440Hz でも行い、サイン波が受信しやすいことを確認したため、本研究ではサイン波を用いることにする。

4.1 複数マイクの受信強度の平均方法

次に、二つのマイクの受信強度の平均により、節を除去する方法を実験する。この場合、二つのマイク間の距離が、波長の半分ほどであれば、平均することで節の影響を除去することが期待できる。しかし、サンプリング周波数の関係上、8kHz が上限である。また、再生側の性能を考え、余裕を見てさらに半分の周波数である 4kHz を用いる。音速を 320m/s とすると、波長は約 8cm となる。ロボットのマイク間の距離は、約 12cm である。結果は、図 3 である。

残念ながら、節の影響を除去し切れていない。左右の信号強度の平均値は、左右単体の信号強度より、単調減少

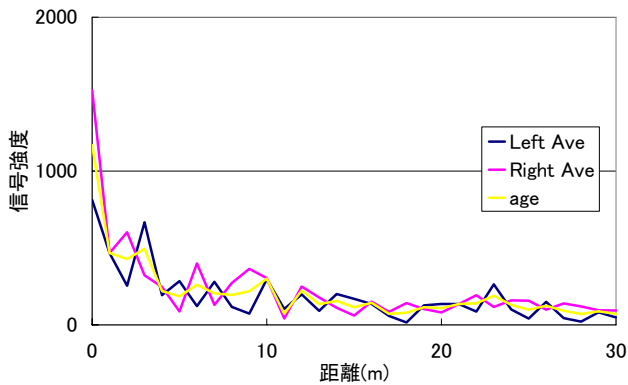


Figure 3: a power signal at 4000Hz.

に近づいているため、距離と再生周波数を適切にすれば、左右単体よりは、有効性がありそうと考えられる。これに関しては、今後の課題とする。

4.2 複数の周波数による距離の推定

次に、複数の周波数を用いることで、節の影響を軽減する方法を実験する。獲得した音データは、サンプリングレート 16kHz の 512 個のデータであるため、フーリエ変換すると 31.25Hz 毎の解析結果が得られる。そこで、8 番目の周波数である 250.00Hz と、9 番目の周波数である 281.25Hz を用い、測定を行った。結果は、図 4 である。

それぞれの節がずれていることが確認できる。250.00Hz の場合は、波長が 1.28m であり、281.25Hz の場合は、1.14m である。両方の波長が理論的に重なるのは、11.24m である。これら二つの周波数の信号強度を組にして、距離を推定するためには、その組から距離が一意に決まる必要がある。それを確認するために、それぞれの信号強度を軸とした二次元空間において、距離を描いた(図 5)。6m までは重ならず二次元平面をたどることができるため、6m までの距離の推定が可能であると考えられる。よって、狭い空間に対する距離の測定は十分可能である。また、二次元にすることで、6m までの重ならない軌跡を確保することができたが、異なる周波数をもう一つ追加し三次元にすることで、さらに重ならない軌跡の確保が期待できる。広い空間における確認は、今後の課題である。

5 おわりに

アドホックネットワークにおける端末間の距離の測定方法として、通常は無線の電波受信強度を用いるが、ここでは音波の信号強度を用いて推定する方法を提案した。端末としてソニーの四足歩行ロボットであるアイボを用い、サイン波の受信感度が十分であることを確認し、低周波側として 250.00Hz と 281.25Hz、そして高周波側として 4000Hz を用いて実験を行った。

ステレオマイクを用いた節の除去は、高周波においても節の影響が現れ、距離に対して単調減少とならないため、

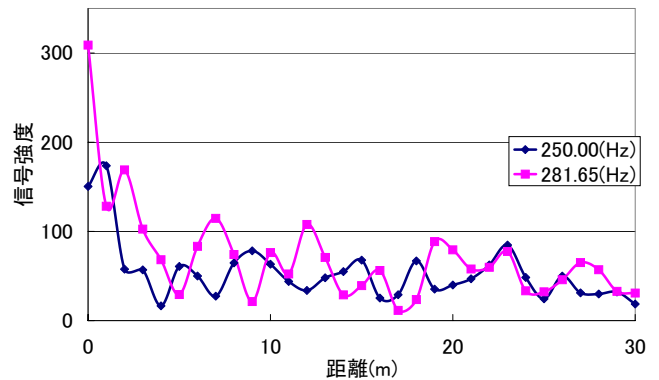


Figure 4: a power signal at 250.00Hz and 281.25Hz.

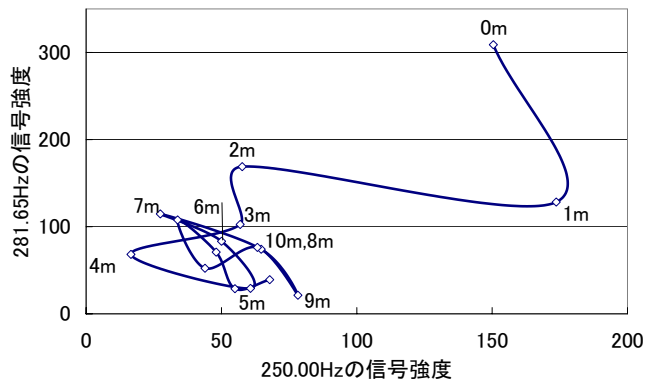


Figure 5: a distance information from 250.00Hz and 281.25Hz.

距離を推定するのは困難であることがわかった。ただし、ロボットを向かい合わせて配置しているため、左右のマイクと音源間の距離には、あまり差がみられない状況である。この差を大きくするためには、使用する音の周波数を高める必要がある。今回は、高周波数は 4000Hz に固定しているため、周波数への依存度は知ることができない。今後の課題として、より適切な周波数を求める必要がある。

複数の周波数を用いた節の除去は、6m ぐらいの狭い空間においては、有効に距離を推定できることがわかった。ただし本実験は、二台のロボットの向きは固定しているため、角度が異なる場合を検討する必要がある。また、用いる周波数の種類を増やすことで、より有効性を高められることがわかった。

今後は、雑音の影響を付加した環境において実験を行い、同様に有効性を示す必要がある。

参考文献

- [1] 村松 潤哉, 牧戸 知史, 鈴木 徳祥: 車々間通信環境での電波伝搬特性に遮蔽車両が与える影響の解析, 信学技法, AP2005-132, pp.13-16, 2005.
- [2] 屋代 智之, Thomas F.La Porta: Nomadic Agent System: インフラに依存しない位置情報サービス提

- 供システム, 情報処理学会論文誌, Vol.46, No.12, pp.2952-2962, 2005.
- [3] 杉山 久佳, 辻岡 哲夫, 村田 正: ネットワーク化された群ロボットによる被災者発見システム, 情報処理学会論文誌, vol. 46, No.7, pp. 1777-1788, 2005.
- [4] 岡本 昌紘, 杉山 久佳, 辻岡 哲夫, 村田 正: 災害救助を目的とした低電力消費型群ロボットネットワークシステム, 信学技報, vol. 105, No. 280, CS2005-21, pp. 31-36, 2005.
- [5] Martijn N.Rooker and Andreas Birk: Combining Exploration and Ad-Hoc Networking in Robocup Rescue, Robocup 2005: Robot Soccer World Cup VIII, Springer, LNAI 3276, pp. 236-242, 2005.
- [6] 板谷 聡子, 長谷川 淳, 長谷川 晃朗, デイビス ピーター, 門脇 直人, 小花 貞夫: 不安定な無線環境における大規模アドホックネットワークの安定化, 情報処理学会論文誌, Vol.46, No.12, pp.2848-2856, 2005.
- [7] 小野 良司, 撫中 達司, 渡辺 尚: アドホックルーティングプロトコルの制御メッセージによる無線 LAN リンクスループットへの影響の分析, 電子情報通信学会論文誌, Vol.J88-A, No.12, pp.1422-1431, 2005.
- [8] T.Clausen and P.Jacquet: Optimized link state routing protocol(OLST), Internet Engineering Task Force(IETF), Internet-Draft: draft-ietf-manet-olsr-11.txt (July, 2003).
- [9] R.Ogier, M.Lewis, and F.Templin, Topology dissemination based on reverse-path forwarding(TBRPF), Internet Engineering Task Force(IETF), Internet-Draft: draft-ietf-manet-tbrpf-11.txt (Oct, 2003).
- [10] C.Perkins, E.Belding-Royer, and S.Das, Ad hoc on-demand distance vector(AODV) routing, Internet Engineering Task Force(IETF), Internet-Draft: draft-ietf-manet-aodv-13.txt (July, 2004).
- [11] D.B.Johnson, D.A.Maltz, Y.Hu, The dynamic source routing protocol for mobile ad hoc networks(DSR), Internet Engineering Task Force(IETF), Internet-Draft: draft-ietf-manet-dsr-10.txt (July, 2004).
- [12] 杉山 久佳, 辻岡 哲夫, 村田 正: アドホックネットワークと自律的パス修復を用いた被災者発見システムによる広域探索, 計測自動制御学会論文集, vol. 41, No.2, pp. 183-185, 2005.
- [13] 滝沢 泰久, デイビス ピーター, 岩井 誠人, 川合 誠, 小花 貞夫: 無線アドホックネットワークによる自律的端末位置推定方式とその特性, 情報処理学会論文誌, Vol.46, No.12, pp.2903-2914, 2005.
- [14] シルビア ニッテル, グァン ジン, 白石 陽: センサネットワークにおけるネットワーク内空間推定, 電子情報通信学会論文誌, Vol.J88-A, No.12, pp.1413-1421, 2005.
- [15] 菊池 祥平, 辻 宏之, 三浦 龍, 佐野 昭: マルチパス環境における周辺散乱モデルに基づく無線局の位置特定法, 電子情報通信学会論文誌, Vol.J87-B, No.12, pp.2020-2028, 2004.
- [16] 筒井 稔, 小長谷 重雄, 香川 忠興: 周波数分散特性を持ったパルス電磁波の到来方位測定法, 電子情報通信学会論文誌, Vol.J89-B, No.1, pp.1-9, 2006.
- [17] R.O.Schmidt: Multiple Emitter Location and Signal Parameter Estimation, IEEE Trans. on Antennas and Propagation, Vol. AP-34, No.3, pp. 276-280, 1986.
- [18] 青島 伸一, 吉澤 信幸, 藪田 哲郎, 羽成 賢一: 超音波センサの距離と感度情報を用いたニューラル型対象物判別システム, 計測自動制御学会論文集, Vol.38, No.1, pp.111-113, 2002.
- [19] 南 正輝, 森川 博之, 青山 友紀: 超音波を用いた自律分散型位置検出システムの設計と実装, 電子情報通信学会論文誌, Vol.J88-A, No.12, pp.1432-1441, 2005.
- [20] 佐々木 公男, 平田 克己: 3点検出による近方場定常不規則音源の3次元位置推定, 計測自動制御学会論文集, Vol.34, No.10, pp.1392-1337, 1998.
- [21] 岡田 徳次, 高橋 真之, 佐藤 秀幸: マイクロホン感度の単一指向性を利用する簡易型3次元音源定位, 計測自動制御学会論文集, Vol.39, No.5, pp.441-448, 2003.
- [22] <http://openr.aibo.com/>
- [23] <http://www.robocup2006.org/sixcms/detail.php?id=56&lang=en>
- [24] <http://www.tzi.de/4legged/bin/view/Website/WebHome>
- [25] OPEN-R プログラミング SIG 著, ソニー株式会社エンタテインメントロボットカンパニー監修: C++でAIBOを自在に動かす - OPEN-R プログラミング入門 -, 株式会社インプレス, 2002.

マルチ全方位ビジョンシステムによる物体認識および自己位置同定手法

Object Recognition and Self-Localization Method

Used Multiple Omnidirectional Vision System

井戸大介, 前田陽一郎

Daisuke IDOU, Yoichiro MAEDA

福井大学大学院

Graduate School of Engineering, University of Fukui

{didou, maeda}@ir.his.fukui-u.ac.jp

Abstract

We have already developed Multiple Omnidirectional Vision System (MOVIS) with three omnidirectional cameras and its calculation method for the measurement of the object position and the self-localization in an autonomous mobile robot. In this research, we propose the efficient recognition method for multiple objects and high-precision self-localization method in an autonomous mobile robot with MOVIS. In this paper, we also report some experimental results to confirm the efficiency of our proposed method by using a RoboCup soccer robot.

1 はじめに

近年、複雑な環境下において適応的に行動する自律移動ロボットに関する研究が盛んに行われている。しかし、自律移動ロボットが複雑な環境情報を獲得し、人間のように柔軟で適応性のある行動を取ることは容易ではない。

RoboCup 中型ロボットリーグのような複数の自律移動ロボットが存在する環境下においては、全方位カメラによ

る全周囲の情報を用いた研究が数多く行われている[1]。しかし、一般に単眼で全方位カメラを用いた場合、計測対象物の高さが既知でない限り正確な距離を測定することは不可能である。また、2台の全方位カメラを垂直に用いた距離計測方法もいくつか提案されているが、これらの方法ではミラー上で圧縮率の異なる情報を基にステレオ計測を行っているため、高精度に距離を計測することは困難である[2, 3, 4]。これに対し本研究室では3台の全方位カメラを用いたマルチ全方位ビジョンシステム MOVIS(Multiple Omnidirectional Vision System) を提案している[5, 6, 7, 8]。本システムでは360度全周にわたり、計測物体の距離と方位を比較的精度良く取得することが可能となる。本論文では、まず MOVIS の基本的なシステム構成について説明する。次に、MOVIS における効率的な異形状物体(サッカーゴール、コーナーボール等)の認識手法について提案する。さらに、MOVIS の高精度な距離計測を生かした自己位置同定手法について提案する。ここでは RoboCup 中型ロボットリーグ規格のサッカーロボット (Figure 1 参照) を用いて計測を行った。また、提案した手法の有効性検証実験を行ったので、その実験結果についても報告する。



Figure 1: MOVIS を搭載したサッカーロボット



Figure 2: MOVIS の概観

2 MOVISの基本構成

2.1 MOVISの原理

筆者らの研究室では、マルチ全方位ビジョンシステム MOVIS をすでに提案している。ここでは、MOVIS の概要について述べる。

MOVIS において、全方位カメラは3台使用し (Figure 2 参照)、それぞれ M_1 、 M_2 、 M_3 とする。本システムは Figure 3 に示すように、移動ロボット上に床面と水平に正三角形に配置する。この時、ロボットの重心と3つの全方位ビジョンが作る正三角形の重心は垂直方向に対し同一点に存在するようにする。正三角形の重心より各頂点を延長した境界線により、各2台のカメラの画像取得範囲を120度ずつ3等分した領域に分割し、各領域をそれぞれ2台のカメラ対 (Figure 3 の場合、 M_1 と M_2 、 M_2 と M_3 、 M_3 と M_1) を用いてステレオ視による三角測量を行なう。

2.2 ロボット座標系における物体位置計測

Figure 3 において位置計測は計測物体 $A(x_a, y_a)$ を全方位ビジョン M_1 、 M_2 を用いて三角測量を行い、物体位置をロボット重心を中心とした座標系で求める。x 軸に水平にある $\overline{M_1M_2}$ に対し、 M_1 における物体 A の傾きは $\theta_1 - \frac{\pi}{6}$ 、 M_2 における物体 A の傾きは $\theta_2 - \frac{5\pi}{6}$ であり、ロボットの重心からビジョンまでの距離を L とする。ビジョン M_1 、 M_2 の中心座標は既知であるため、ロボット座標系における物体 A の位置 (x_a, y_a) は、(1)、(2) 式で求められる。

$$x_a = \frac{\sqrt{3}}{2}L \cdot \frac{\tan(\theta_2 - \frac{5\pi}{6}) + \tan(\theta_1 - \frac{\pi}{6})}{\tan(\theta_2 - \frac{5\pi}{6}) - \tan(\theta_1 - \frac{\pi}{6})} \quad (1)$$

$$y_a = \frac{1}{2}L + \sqrt{3}L \cdot \frac{\tan(\theta_2 - \frac{5\pi}{6}) \cdot \tan(\theta_1 - \frac{\pi}{6})}{\tan(\theta_2 - \frac{5\pi}{6}) - \tan(\theta_1 - \frac{\pi}{6})} \quad (2)$$

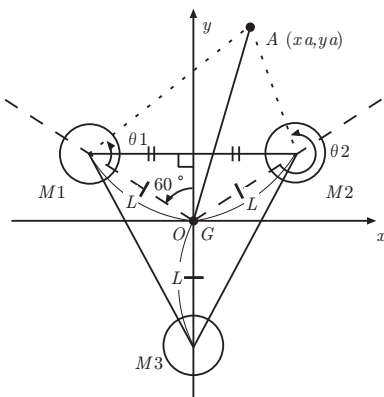


Figure 3: ロボット座標系における物体位置計測

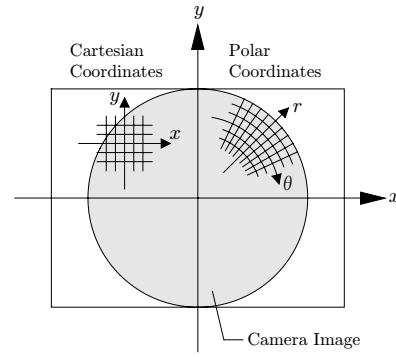


Figure 4: 画像の極座標表現

3 MOVISによる物体認識手法

MOVIS において距離計測は原則的に各カメラの物体への角度 θ_n のみに基づき行われる。全方位ビジョンは下向きに置かれた双曲面ミラーと上向きにカメラが設置された装置であり、実際の距離計測に必要な情報は双曲面ミラー上のみ存在する事になる。移動ロボットによる物体認識において高性能な画像処理装置の搭載は困難なため、本研究では MOVIS と全方位ビジョンの特長を生かした効率的な物体認識手法について提案する。

まず全方位カメラにおいて、Figure 4 のように画像上の座標 (x, y) を双曲面ミラーの中心 $O(Ox, Oy)$ を中心とする以下のような極座標系 (θ, r) で表現する。

$$x = \cos(\theta) \cdot r \quad (3)$$

$$y = \sin(\theta) \cdot r \quad (4)$$

θ に対して r 方向に注目している色塊 (RoboCup 中型ロボットリーグの場合、黄色、青色、緑色) のピクセル数をカウントすることにより、Figure 5 のような簡易的な全周の展開画像 (ヒストグラム) が得られる。

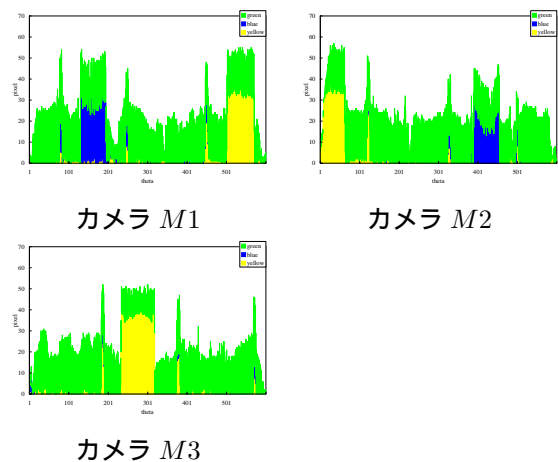


Figure 5: 各カメラにおけるヒストグラム

このヒストグラムにおける色塊の重心位置より、直接各カメラの物体への θ_n が得られる。MOVIS はこの θ_n のみで物体の距離と方位を正確に計測できるので、複雑な画像処理等を省く事ができ、高速に全周の状況を把握する事ができる (Figure 6 参照)。

4 MOVIS による自己位置同定手法

4.1 自己位置同定の原理

RoboCup 中型ロボトリグにおいて自己位置同定は重要な課題である。自己位置同定手法は過去に多くの提案がなされているが実用的なものはほとんど存在しない。ロボットの置かれている様々な状況によって最適な自己位置同定手法は一般的に変化する。そこで、MOVIS の正確な距離計測を生かして、現在のロボットの置かれている状況から適応的に手法を使い分ける自己位置同定手法を提案する。本手法では黄色と青色のゴールの両端エッジおよび中心位置をランドマークとして使用した。

本研究で自己位置同定手法は以下のものを用い、二つのランドマークの選択も動的に行う。

1. 二つのランドマークの絶対座標と相対座標を用いた同次変換による手法 [手法 1](Figure 7 参照)
2. 二つのランドマークの相対角度とどちらか一方への相対距離を用いた手法 [手法 2](Figure 7 参照)
3. 二つのランドマークとロボットが成す三角形の外接円の交点を用いた手法 [手法 3](Figure 7 参照)

手法 1 において、自己位置 $R(X_r, Y_r)$ は (5), (6) 式で計算できる。

$$X_r = \sin \beta y_p - \cos \beta x_p \quad (5)$$

$$Y_r = Fd - \sin \beta x_p - \cos \beta y_p \quad (6)$$

ここで、

$$\beta = \arctan \frac{x_p - x_q}{y_p - y_q} \quad (7)$$

手法 2 において、自己位置 $R(X_r, Y_r)$ は (8), (9) 式で計算できる。

$$X_r = d \cdot \sin \left(\pi - \left(\theta + \arcsin \left(\frac{\sin \theta \cdot d}{2Fd} \right) \right) \right) \quad (8)$$

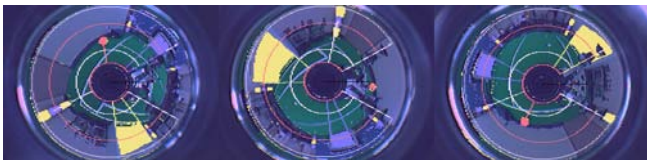


Figure 6: MOVIS による物体認識の様子

$$Y_r = Fd - d \cdot \cos \left(\pi - \left(\theta + \arcsin \left(\frac{\sin \theta \cdot d}{2Fd} \right) \right) \right) \quad (9)$$

手法 3 において、自己位置 $R(X_r, Y_r)$ は (10) ~ (12) 式で計算できる。

$$X_r = +\sqrt{r_p^2 - (Y_r - Y_{cp})^2} \quad (x_{p1}^2 + y_{p1}^2 < x_{p2}^2 + y_{p2}^2) \quad (10)$$

$$X_r = -\sqrt{r_p^2 - (Y_r - Y_{cp})^2} \quad (x_{p1}^2 + y_{p1}^2 > x_{p2}^2 + y_{p2}^2) \quad (11)$$

$$Y_r = \frac{(r_p^2 - Y_{cp}^2) - (r_q^2 - Y_{cp}^2)}{2 \cdot (Y_{cq} - Y_{cp})} \quad (12)$$

4.2 自己位置同定実験

提案手法の有効性を示すため実証実験を行った。フィールドの任意の座標にロボットを置き、一定速度で回転しながら 100 回計測を行った。実験条件は Table 1 のとおりである。

Table 1: 実験条件

フィールドサイズ		400cm × 350cm
ロボットサイズ		約 50cm × 50cm
搭載パソコン	本体	DELL LATITUDE
	CPU	celeron 1.5GHz
	メモリ	128MB
カメラ	解像度	320 × 240
	フレームレート	30fps

実験結果を Figure 8 ~ 16 に示す。これらの実験結果より、かなり高い精度で自己位置が同定できている事を確認した。しかし Figure 8 ~ 10 において、一様に黄色ゴールの方向へ同定位置がずれていることが分かる。この原因として、青色ゴールの計測に誤差が生じている (具体的に

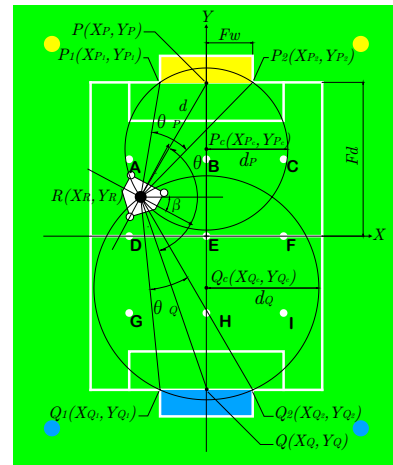


Figure 7: 自己位置同定手法

は青色ゴールを実際よりも遠く認識している) ことが考えられる。これは YUV 表色系において青色を抽出するための閾値範囲の調整が困難であるため青色ゴールの認識に支障が出ていると考えられる。MOVIS の性質上、各カメラにおける対象物体への角度 θ_n を精度よく計測しなければならないが、青色ゴールの θ_n の精度が悪くなり結果として自己位置同定に誤差が生じてしまったものと考えられる。

5 おわりに

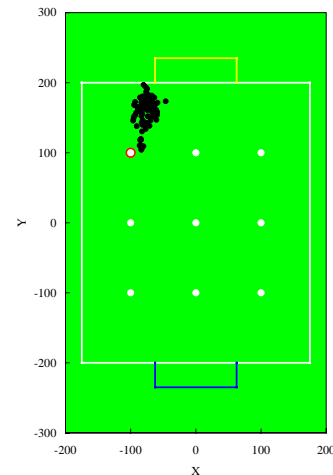
本論文ではマルチ全方位ビジョンシステム (MOVIS) による物体認識手法および自己位置同定手法について提案した。さらに検証実験により、提案手法の有効性が検証できた。

しかしながら、場所によっては誤差が生じることも分かったためより複雑な実験環境でも高精度な自己位置同定ができるよう、MOVIS の更なる改良が必要とされる。今後の課題として MOVIS によるコーナーポールの選択的認識、類似複数物体 (敵、味方ロボット) の認識および位置計測などが考えられる。

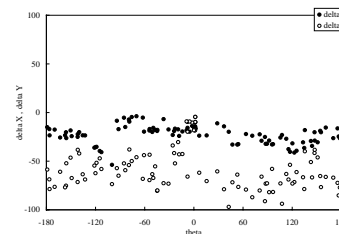
参考文献

- [1] 知原 伸悟, 川端 邦明, 浅間 一, 三島 健稔: “全方位視覚センサによる移動ロボットの自己位置推定,” 第 20 回日本ロボット学会学術講演予稿集, CD-ROM, 3A25 (2002)
- [2] J.Gluckman, K.Nayar and J.Thoresz: “Real-Time Omnidirectional and Panoramic Stereo,” *Proc. of Image Understanding Workshop*, Vol.1, pp.299-303 (1998)
- [3] 子安 大士, 三浦 純, 白井 良明: “実時間全方位ステレオを用いたロボットナビゲーションのための移動物体認識,” 2001 年ロボティクス・メカトロニクス講演会予稿集, CD-ROM (2001)
- [4] T.Matsuoka, M.Araoka, T.Hasegawa and et. al: “Localization and Obstacle Detection Using Omnidirectional Vertical Stereo Vision,” *RoboCup 2001*, LNAI2377, pp.429-434 (2002)
- [5] W.Shimizuhira and Y.Maeda: “Self-Localization Method Used Multiple Omnidirectional Vision System,” *SICE Annual Conference 2003*, CD-ROM, pp.2796-2799 (2003)
- [6] 清水平 渉, 前田 陽一郎: “マルチ全方位ビジョン MOVIS を用いた物体位置計測および自己位置同定手法,” 第 21 回日本ロボット学会学術講演会予稿集, CD-ROM, 3K1A (2003)

- [7] 清水平 渉, 前田 陽一郎: “マルチ全方位ビジョン MOVIS を搭載したサッカーロボットの階層型ファジィ行動制御,” 第 22 回日本ロボット学会学術講演会予稿集, CD-ROM, 2F16 (2004)
- [8] 井戸 大介, 清水平 渉, 前田 陽一郎: “MOVIS を搭載した自律移動ロボットによる物体位置計測および誤差補正手法,” 第 23 回日本ロボット学会学術講演会, CD-ROM, 1B15 (2005)
- [9] 津崎 亮一, 吉田 和夫: “ファジィポテンシャル法に基づく全方位視覚を用いた自律移動ロボットの行動制御,” *日本ロボット学会誌*, Vol.21, No.6, pp.656-662 (2003)
- [10] Y.Takahashi, K.Hikita, and M.Asada: “Incremental Purposive Behavior Acquisition based on Self-Interpretation of Instructions by Coach,” *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.686-693 (2003)
- [11] 中村 恭之, 高橋 泰岳編著: 中型ロボットの基礎技術, 共立出版 (2005)

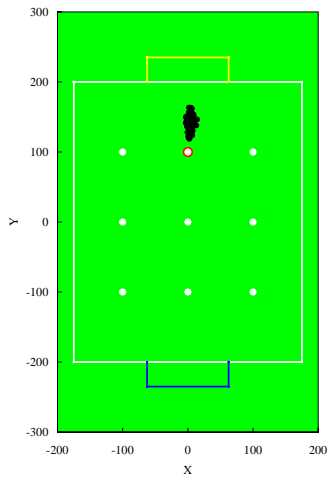


(a) 同定位置

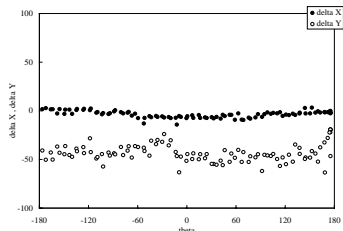


(b) 位置誤差

Figure 8: 自己位置同定結果 (-100cm,100cm)

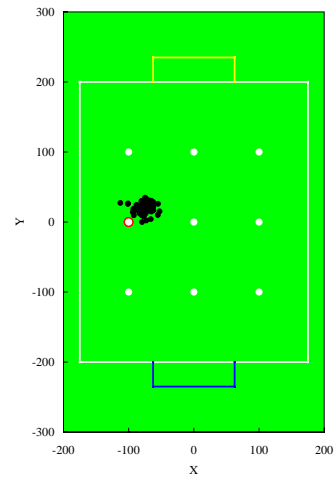


(a) 同定位置

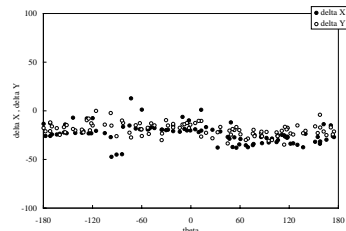


(b) 位置誤差

Figure 9: 自己位置同定結果 (0cm,100cm)

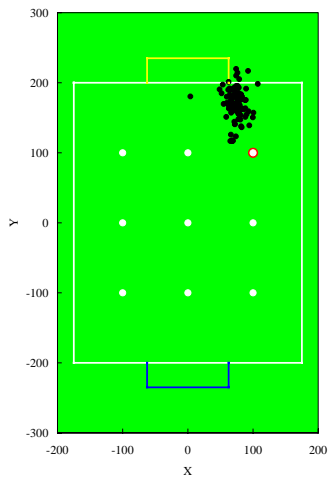


(a) 同定位置

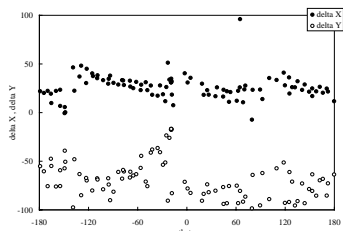


(b) 位置誤差

Figure 11: 自己位置同定結果 (-100cm,0cm)

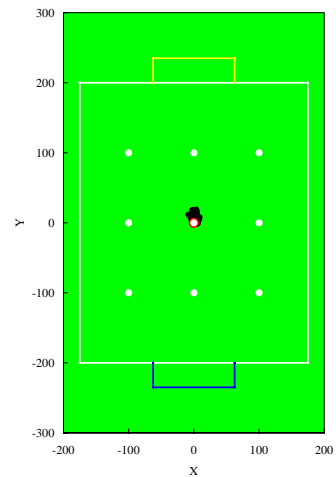


(a) 同定位置

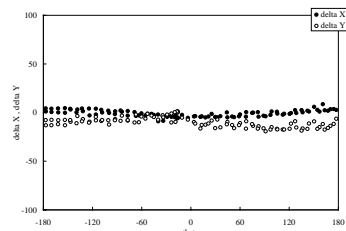


(b) 位置誤差

Figure 10: 自己位置同定結果 (100cm,100cm)

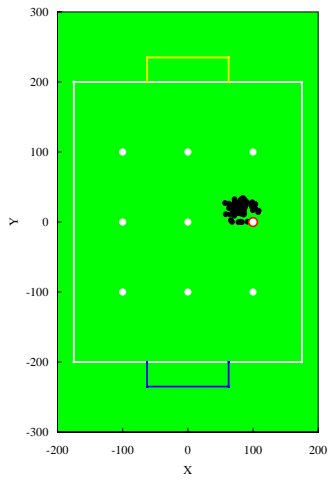


(a) 同定位置

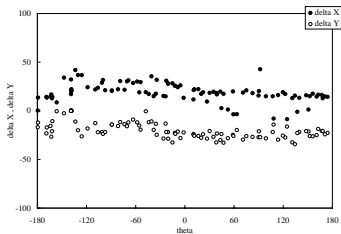


(b) 位置誤差

Figure 12: 自己位置同定結果 (0cm,0cm)

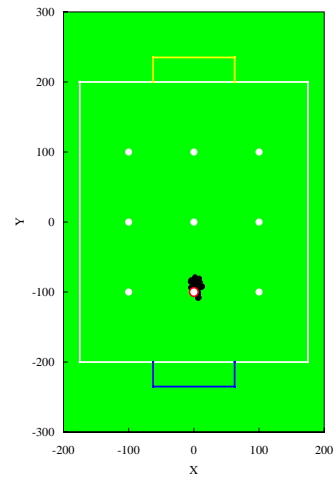


(a) 同定位置

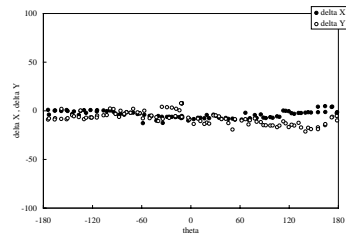


(b) 位置誤差

Figure 13: 自己位置同定結果 (100cm,0cm)

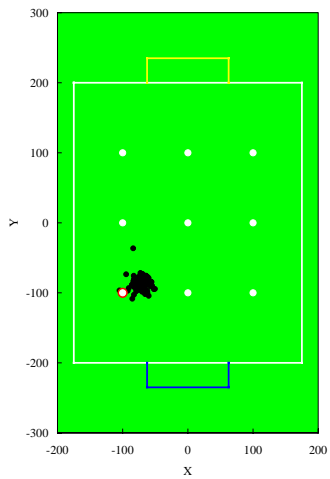


(a) 同定位置

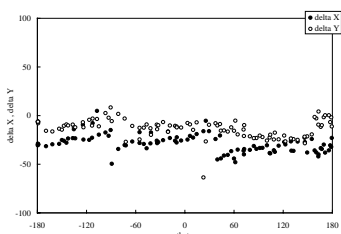


(b) 位置誤差

Figure 15: 自己位置同定結果 (0cm,-100cm)

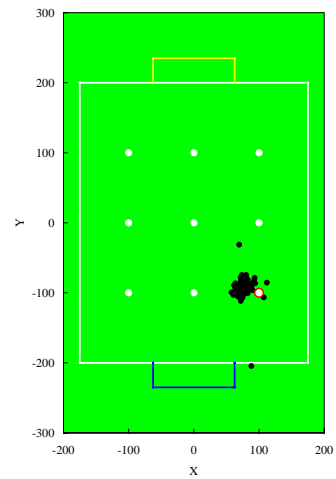


(a) 同定位置

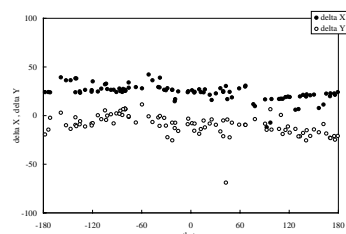


(b) 位置誤差

Figure 14: 自己位置同定結果 (-100cm,-100cm)



(a) 同定位置



(b) 位置誤差

Figure 16: 自己位置同定結果 (100cm,-100cm)

Inferring other's intention based on estimated state value of self

Yasutake Takahashi, Teruyasu Kawamata, Tom Tamura, and Minoru Asada

Dept. of Adaptive Machine Systems,
Graduate School of Engineering, Osaka University
Handai FRC
Yamadaoka 2-1, Suita, Osaka, 565-0871, Japan
{yasutake, kawamata, tamura, asada}@er.ams.eng.osaka-u.ac.jp

Abstract

Recognition of other agent intention in a multi-agent environment is a very important issue to realize social activities, for example, imitation learning, understanding intention, cooperative/competitive behavior, and so on. Conventional approaches to infer the other agent intention need a precise trajectory in Cartesian or joint space that is sometimes hard to measure from the viewpoint of an observer. It is also difficult to estimate a same intention but with different realizations because they try to match just a certain trajectory during the trial. We propose a novel method of inference of other agent's intention based on state value estimation. The method does not need a precise world model or coordination transformation system to deal with view dependency. This paper shows an observer can infer an intention of other not by precise object trajectory in Cartesian space but by estimated state value transition during the observed behavior.

1 Introduction

Inference of others' intentions what they like to do is one of the most formidable issues in multi-agent systems in which actions appropriate for the others' intentions are needed to accomplish the cooperative tasks. For example Schaal et al. [4] proposed a motor learning method through imitation of teacher's behaviors. They assume that a learner can observe all state variables and their trajectories in Cartesian coordinate system of the environment or the joint space of the others and the learner imitates manipulative tasks or gestures. Doya et al. [2] proposed to estimate intention of other agent for imitation learning and/or cooperative behavior acquisition based on multi-module learning system. Takahashi et al. [6] proposed a method that interprets instruction given by a coach and divides the given complicated task to a number of simple sub-tasks each of which can be learned

with a simple behavior learning module with limited capability. Most existing approaches assume the detailed knowledge of the task, the environment, and the others (their body structure and sensor/actuator configuration) based on which they can transform the observed sensory data of the others' behaviors into the Cartesian coordinate system of the environment or the joint space of the others to infer their intentions. However, such an assumption seems unrealistic in the real world and brittle to the sensor/actuator noise(s) or any possible changes in the parameters. In other words, it is very difficult to infer others' intentions based only on these geometric parameters.

On the other hand, another approach that estimates behavior of others through observer's viewpoint without any coordination conversion has been proposed, too. Ledezma et al. [3] proposed to make a classifier to label other agent's behavior based on observation and use this classifier to label the behavior. Their method, however, needs a full teaching data of a set of labels and sequence observation in order to model the other agent actions and cannot handle the change the sequence of the other agent's actions even if it does the same task. Takahashi et al. [5] presented an approach that constructs a set of state transition models for the opponent behaviors from a viewpoint of observer and selects an appropriate behavior for observer according to a current situation in which one of the models matches. The observer can choose one model according to the other agent's behavior, however, it cannot infer the intention of the other agent.

Recently, reinforcement learning has been studied well for motor skill learning and robot behavior acquisition. It generates not only an appropriate policy (map from states to actions) to achieve a given task but also an estimated discounted sum of reward value that will be received in future while the robot is taking the optimal policy. We call this estimated discounted sum of reward "state value." This state value roughly indicates closeness to a goal state of the given task, that is, if the agent is getting closer to the goal, the state value becomes higher. This suggests that the observer may understand which goal the agent likes to achieve if the state value of

the corresponding task is going higher.

The relationship between an agent and objects such that the agent gets close to the object or the agent faces to a direction is much easier to understand from the observation, and therefore such qualitative information should be utilized to infer what the observed agent likes to do. The information might be far from precise ones, however, it keeps topological information and we can acquire good estimation of temporal difference of state value with this method.

Then, we propose a novel method to apply the above idea to infer the others' intentions supposing that the observer has already estimated the state values of all kinds of tasks the observed agent can do. The method does not need a precise world model or an accurate coordination transformation system to cope with the problem of view dependency. We apply the method to a simple RoboCup situation where the agent has kinds of tasks such as navigation, shooting a ball into a goal, passing a ball to a teammate, and so on, and the observer judges which task the agent is now achieving from the observation with estimated state values. The preliminary experiments are shown and future issues are discussed.

2 Intention Inference by State Value Estimation

In this section, a rough description of state value function and behavior inference is described. We assume that the observer has already acquired a number of behaviors based on a reinforcement learning method. Each behavior module can estimate state value at arbitrary time t to accomplish the specified task. Then, the observer watches the performer's behavior and maps the sensory information from an observer viewpoint to the agent's one with a mapping of state variables. The behavior modules estimate the state value of the observed behavior and the system selects ones that matches estimation of state value.

2.1 State Value Function

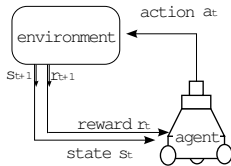


Figure 1: A basic model of agent-environment interaction

Fig.1 shows a basic model of reinforcement learning. An agent can discriminate a set S of distinct world states. The world is modeled as a Markov process, making stochastic transitions based on its current state and the action taken by the agent based on a policy π . The agent receives reward r_t at each step t . State Value V^π , discounted sum of the reward received over time under

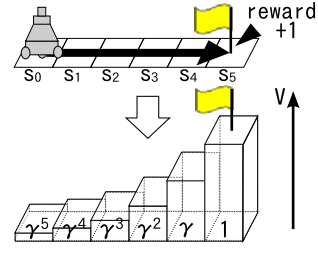


Figure 2: Sketch of state value propagation

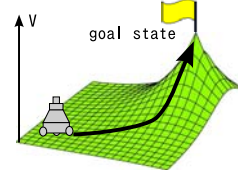


Figure 3: Sketch of a state value function

execution of policy π , will be calculated as follows:

$$V(s) = \sum_{t=0}^{\infty} \gamma^t r_t . \quad (1)$$

Figs.2 and 3 show sketches of a state value function where a robot receives a positive reward when it stays at a specified goal while zero reward else. The state value will be highest at the state where the agent receives a reward and discounted value is propagated to the neighbors states (Fig.2). As a result, the state value function seems to be a mountain as shown in Fig.3. The state value becomes bigger and bigger if the agent follows the policy π .

2.2 Basic Idea of Intention Recognition

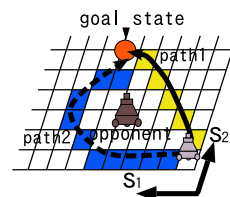


Figure 4: Sketch of different behaviors in a grid world

Fig.4 shows an example task of navigation in a grid world. There is a goal state at the top center of the world. An agent can move one of the neighbor grids every one-step. It receives a positive reward only when it stays at the goal state while zero else. There are various optimal policies for this task as shown in Fig.4. If one tries to match the action that the agent took and the one based on a certain policy in order to infer the agent's intention, you have to maintain various optimal policies and evaluate all of them in the worst case.

On the other hand, if the agent follows an optimal policy, the state value is going up even if the agent takes

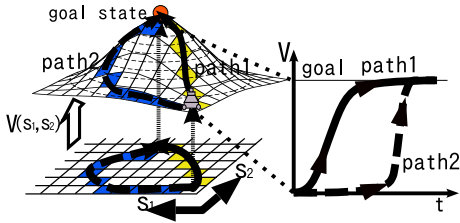


Figure 5: Inferring intention by the change of state value

an arbitrary policy from the optimal ones. Fig.5 shows that the state value becomes larger even if the agent takes different paths. We can regard that the agent takes an action based on one policy when the state value is going up even if it follows various kind of policies.

This indicates a possibility of robust intention recognition even if they would be several optimal policies for the current task. An agent tends to acquire various policies depending on the experience during learning. The observer cannot practically estimate the agent's experience beforehand, therefore, it needs a robust intention recognition method provided by the estimation of state values.

2.3 Modular Learning System

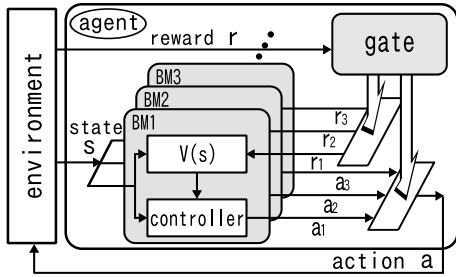


Figure 6: Modular Learning System

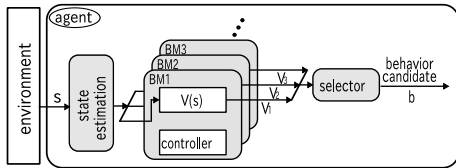


Figure 7: Behavior inference diagram

In order to evaluate a number of behaviors simultaneously, we adopt a modular learning system. Jacobs and Jordan [1] proposed a mixture of experts, in which a set of the expert modules learn and the gating system weights the output of each expert module for the final system output. Fig.6 shows a sketch of such a modular learning system. We prepare a number of behavior modules each of which acquired a state value function for one goal-oriented behavior. A learning module has a controller that calculates an optimal policy based on the

state value function. Gating module selects one output from a module according to the agent's intention.

2.4 Intention Inference under Multiple Candidates

At intention inference stage, the system uses same behavior modules as shown in Fig.7. While an observer watches an behavior of a performer, the system estimates the relationship between the agent and objects such as rough direction and distance of the objects from the agent. Then, each behavior module estimates the state value based on the rough estimated state of the agent and sends it to the selector. The selector watches the sequence of the state values and selects a set of possible behavior modules of which state values are going up as the performer is taking the behavior. As mentioned in 2.1, if the state value goes up during a behavior, it means the module seems valid for explaining the executing behavior execution. The goal state/reward model of this behavior module represents the intention of the agent.

Here we define reliability g that indicates how much the intention inference would be reasonable for the observer as follow:

$$g = \begin{cases} g + \beta & \text{if } V(s_t) - V(s_{t-1}) > 0 \text{ and } g < 1 \\ g & \text{if } V(s_t) - V(s_{t-1}) = 0 \\ g - \beta & \text{if } V(s_t) - V(s_{t-1}) < 0 \text{ and } g > 0 \end{cases}$$

where β is an update parameter, which is 0.1 in this paper. This equation indicates that the reliability g will become large if the estimated state value rises up and it will become low when the estimated state value goes down. We put another condition in order to keep g value from 0 to 1.

3 Task and Environment

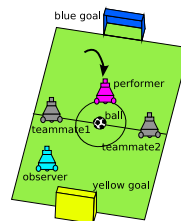


Figure 8: Environment



Figure 9: A real robot

Fig.8 shows a situation the agents are supposed to encounter. An agent shows a behavior and the observer estimates the behavior using a set of behavior modules of its own. Fig.9 shows a mobile robot we have designed and built. Fig.10 shows the viewer of our simulator for our robots and the environment. The robot has a normal perspective camera in front of its body. It has an omni-directional camera, however, it doesn't use it, here. A simple color image processing is applied to detect the ball, the interceptor, and the receivers on the image in

Table 1: Prepared Modules and their state variables

Module	State variables
GoToBall	ball position y on the image of perspective camera
GoToYellow	yellow goal position y on the image of perspective camera
GoToBlue	blue goal position y on the image of perspective camera
ShootYellow	ball position y , yellow goal position y , and angle between them θ on the image
ShootBlue	ball position y , blue goal position y , and angle between them θ on the image
PassToTeammate1	ball position y , teammate 1 position y , and angle between them θ on the image
PassToTeammate2	ball position y , teammate 2 position y , and angle between them θ on the image



Figure 10: Viewer of simulator

real-time (every 33ms). The left of Fig.10 shows a situation the agent encounters while the top right images show the simulated ones of the normal and the bottom right omni vision systems. The mobile platform is an omni-directional vehicle (any translation and rotation on the plane). Table 1 shows a list of prepared behavior modules and their state variables. The observer has learned the behaviors and its state value estimator based on a reinforcement learning method beforehand.

3.1 State Variables and Estimation

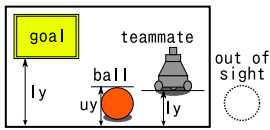


Figure 11: State variables representing distances to the objects

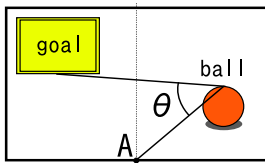


Figure 12: A state variable θ representing the positional relationship between the objects

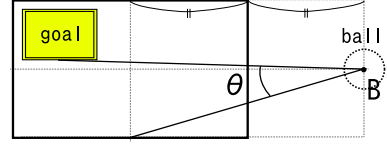


Figure 13: A state variable θ when one of the objects is out of sight

We use the distances of the ball, goal, and player from the agent and their relative angles between them on the image of the frontal camera on the robot. Figs.11 and 12 show examples of those state variables. We divide this state space into a set of region to obtain state id. The space of position value is quantized into 6 subspaces and the space of relative angle between objects into 5 spaces here. Behavior modules define their policy and state value function in this state space.

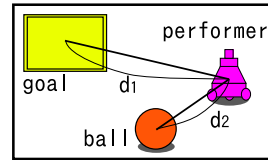


Figure 14: Estimated state variables representing distances

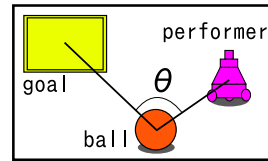


Figure 15: An estimated state variable θ representing the position relation among objects

When an observer infers an intention of the performer, it has to estimate its state. Here, we introduce a simple method of state estimation of the performer. Fig.14 shows the estimated distances from the agent and the objects. Fig.15 shows the estimated angle between the objects. The observer uses these estimated state for estimation of state value instead of its own state shown in Figs.11 and 12. These estimated states with this method

are far from precise ones, however, it keeps topological information and we can acquire good estimation of temporal difference of state value with this method.

4 Experiments

The observer has learned a number of behaviors shown in Table 1 before it tried to infer the performer’s intention. We gave the observer many experiences enough to cover all exploration space in state space.

4.1 Same behavior demonstration

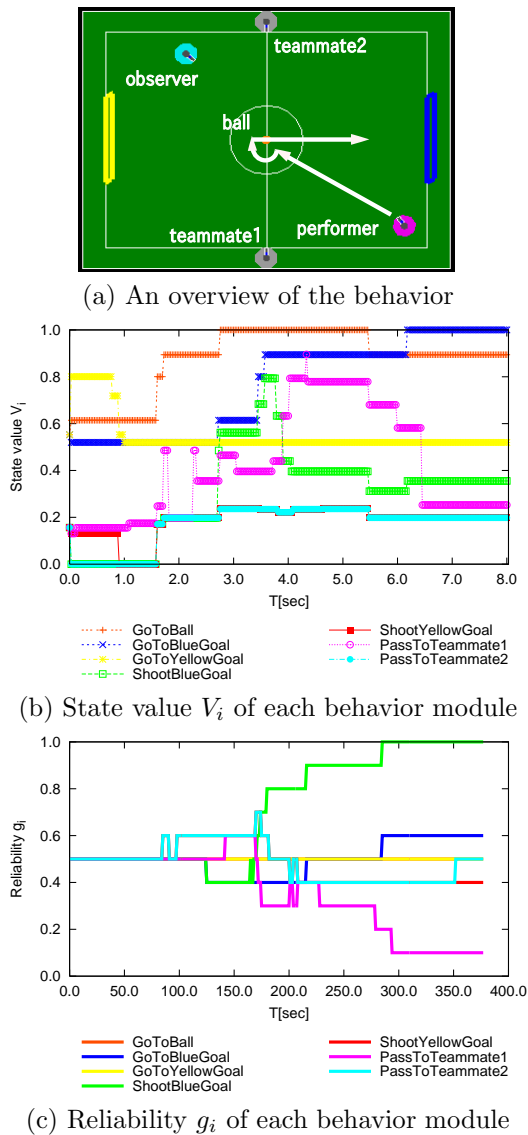


Figure 16: Inferring intention of a performer trying to shoot a ball to the blue goal

After the behavior acquisition, we let the performer play one of the behaviors from Table 1 and the observer infers which behavior the other is taking. Fig.16 shows an example behavior performed by another agent. The performer showed exactly same behavior that the observer acquired in behavior learning stage, here. The bottom right agent shows "ShootBlue" behavior and the top left observer watches the behavior. The observer



(a) An overview of the behavior

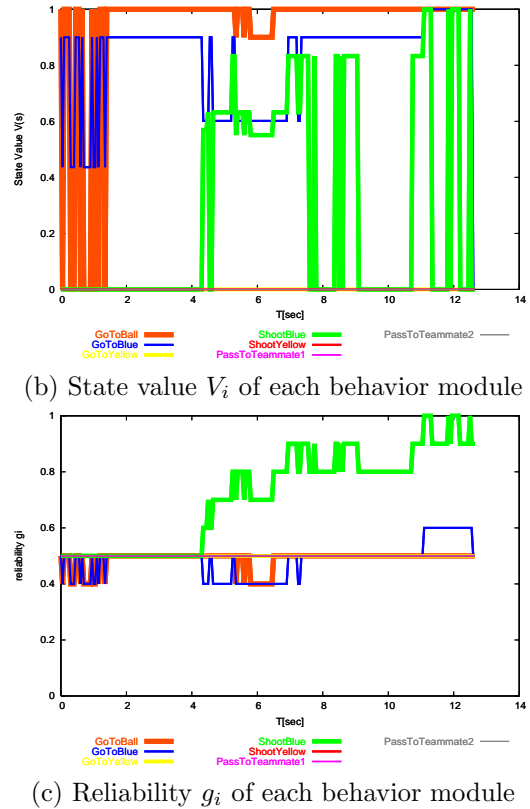
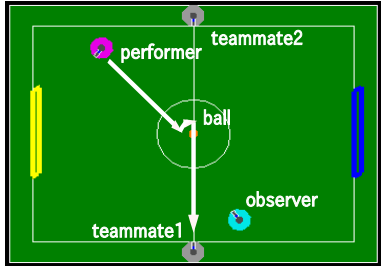


Figure 17: Inferring intention of the performer trying to shoot ball to the blue goal in a real robot experiment

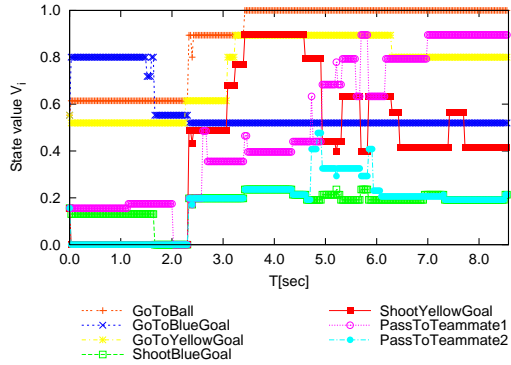
tries to keep the performer in own perspective during the behavior. Fig.16 (a) shows the sequence of the behavior. Fig.16 (b) and (c) show sequences of estimated state value and reliability of the inferred intentions of the agent, respectively. The green line indicates the behavior of shooting a ball into a blue goal and goes up during the trial. The observer successfully inferred the intention of the performer.

Figure 17 shows a result of inferring intention of the performer trying to shoot ball to the blue goal in a real robot experiment. The situation and the result are similar to the simulation and it shows successfully infer the intention of the performer.

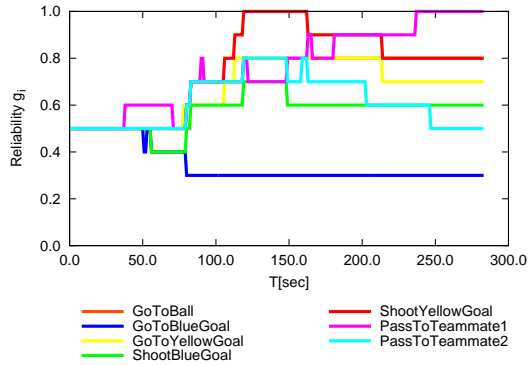
Fig.18 shows an example passing behavior performed by an agent, a sequence of estimated state value of each modules, and a sequence of reliability of inferred intention, respectively. These figures show that the observer



(a) An overview of the behavior



(b) State value V_i of each behavior module



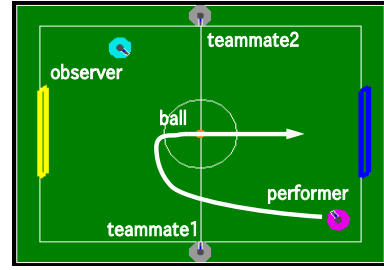
(c) Reliability g_i of each behavior module

Figure 18: Inferring intention of a performer trying to pass a ball to teammate1

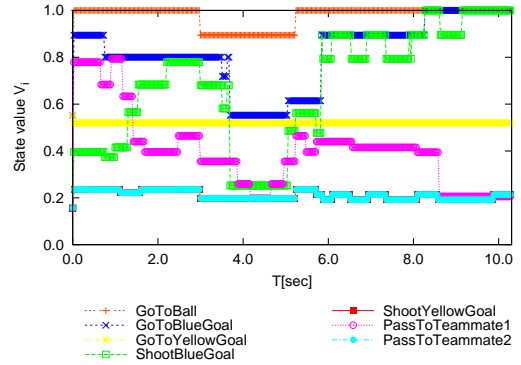
can infer the performer's passing behavior (purple line), too. The orange line indicates the reliability of going to a ball behavior and it also goes up during the trial because the agent is continuously approaching to the ball during the trial to pass it to the teammate.

4.2 Different behavior demonstration

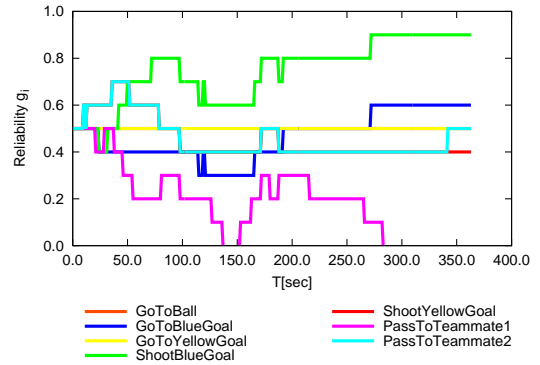
The observer cannot assume that the performer will take exactly same behavior even if its intention is same. We prepare other different behaviors for the demonstration of the performer. Fig.19 (a) shows an example of the different shooting behavior demonstrated by the performer. The learned behavior by the observer is a zippy motion as shown in Fig.16. On the other hand, the demonstrated behavior is a more smooth motion. Therefore, the state transition probability will be different from each other. Figs.19 (b) and 19 (c) show sequences of estimated state value of each modules and a sequence of reliability of inferred intention, respectively. These figures show that



(a) An overview of the behavior



(b) State value V_i of each behavior module



(c) Reliability g_i of each behavior module

Figure 19: Inferring intention of a performer trying to shoot a ball to the blue goal with different manner

the observer can infer the performer's intention of shooting (green line).

4.3 Comparison with System based on Coordination Translation

In this section, we compare performances between our proposed method and the one based on state estimation using coordinate transformation system and tracing state transition probability that is proposed by others, for example [2]. In order to estimate the state value of a behavior module through the observation of the performer, there must be a rough coordinate transformation matrix beforehand. Figs.20 and 21 show a rough sketch of the transformation system. In order to estimate y position on the performer's view image, the observer assumes there are tiles on the floor, maps the positions of an object and the performer, estimates rough distance between them, and maps the distance to the y position on the image of the observer's view. Fig.21 shows a

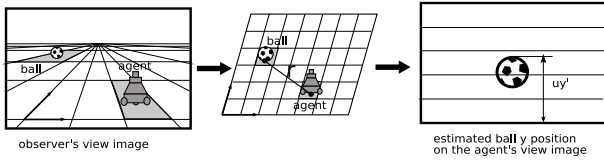


Figure 20: Estimation of y position of the performer's image

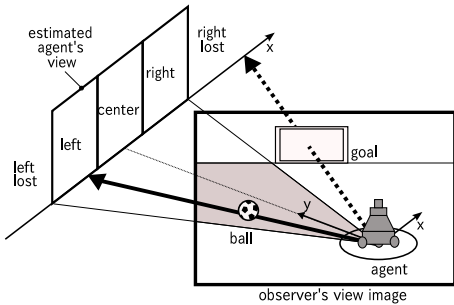


Figure 21: Estimation of direction from a performer to an object

sketch of estimation of direction from the performer to an object. The lower right rectangle shows an example image from the perspective camera and it captures the performer, a ball, and a goal. We assume that it can detect a direction of the performer on the image under a vision system. We put a potential image plane in front of the agent and estimate rough x positions of the objects on the image of the performer's view. Fig.21 shows that the ball is mapped to the left side on the image and the goal to an area of lost to the right side from the camera image. Table 2 shows the success rate

Table 2: Inferring intention performances of the proposed method and the one with coordination transformation system

	Proposed method	Method based on state trans. prob.
ShootBlueGoal1	84%	24%
ShootBlueGoal2	78%	11%
ShootYellowGoal	86%	20%
PassToTeammate1	76%	34%

of intention inference of the proposed method and the one with the coordination transformation system. The proposed method shows much better results over the behaviors than the one with the coordination transformation system. "ShootBlueGoal1" indicates a case of inference of shooting behavior identical to the observer's one. "ShootBlueGoal2" indicates a case of inference of shooting behavior but different from the observer's one.

5 Future work

This basic idea can be applied for not only intention inference but also cooperative behavior acquisition. How to define a reward function for cooperative behavior acquisition in multi-agent system is one of the most interesting issues. The proposed method can infer other's intention and estimate the reward/state value of the agent for each step. This indicates that the observer can explore some actions and evaluate how much they will contribute to the other efficiently. Then, it can learn cooperative behavior based on a certain reinforcement learning approach without any heuristic/hand-coded reward function by which it evaluates a reward of itself based on the estimated reward/state value of the other agent.

References

- [1] R. Jacobs, M. Jordan, Nowlan S, and G. Hinton. Adaptive mixture of local experts. *Neural Computation*, 3:79–87, 1991.
- [2] Doya K., Sugimoto N., Wolpert D.M., and Kawato M. Selecting optimal behaviors based on contexts. In *International Symposium on Emergent Mechanisms of Communication*, pages 19–23, 2003.
- [3] Agapito Ledezma, Ricardo Aler, Araceli Sanchis, and Daniel Borrajo. Predicting opponent actions by observation. In D. Nardi et al., editor, *RoboCup2004*, pages 286–296. Springer-Verlag Berlin Heidelberg, 2005.
- [4] Stefan Schaal, Auke Ijspeert, and Aude Billard. Computational approaches to motor learning by imitation, 2004.
- [5] Yasutake Takahashi, Kazuhiro Edazawa, Kentarou Noma, and Minoru Asada. Simultaneous learning to acquire competitive behaviors in multi-agent system based on modular learning system. In *RoboCup 2005 Symposium papers and team description papers*, pages CD-ROM, Jul 2005.
- [6] Yasutake Takahashi, Tomoki Nishi, and Minoru Asada. Self task decomposition for modular learning system through interpretation of instruction by coach. In *RoboCup 2005 Symposium papers and team description papers*, pages CD-ROM, Jul 2005.

© 2006 Special Interest Group on AI Challenges
Japanese Society for Artificial Intelligence
社団法人 人工知能学会 AI チャレンジ研究会

〒162 東京都新宿区津久戸町 4-7 OSビル 402号室 03-5261-3401 Fax: 03-5261-3402

(本研究会についてのお問い合わせは下記にお願いします.)

AI チャレンジ研究会

主査

奥乃 博

京都大学大学院 情報学研究科

知能情報学専攻 音声メディア分野

〒606-8501 京都市左京区吉田本町

Tel: 075-753-5376 Fax: 075-753-5977

okuno@i.kyoto-u.ac.jp

担当幹事

浅田 稔

大阪大学大学院 工学研究科

知能・機能創成工学専攻 先導的融合工学講座

〒565-0871 大阪府吹田市山田丘 2-1

Tel: 06-6879-7349 Fax: 06-6879-7348 /

JST ERATO 浅田共創知能システムプロジェクト

asada@ams.eng.osaka-u.ac.jp

担当幹事

光永 法明

(株) 国際電気通信基礎技術研究所

知能ロボティクス研究所

〒619-0288 京都府相楽郡精華町光台 2-2-2

Tel: 0774-95-1401 Fax: 0774-95-1408

mitunaga@atr.jp

幹事

中臺 一博

(株) ホンダ・リサーチ・インスティテュート・
ジャパン

〒351-0114 埼玉県和光市本町 8-1

Tel: 048-462-5219 Fax: 048-462-5221 /

東京工業大学大学院 情報理工学研究科

nakadai@jp.honda-ri.com

Executive Committee

Chair

Hiroshi G. Okuno

Dept. of Intelligence Sci. and Tech.,

Graduate School of Informatics,

Kyoto University

Yoshida-honmachi Sakyo-ku,

Kyoto, 606-8501, JAPAN

Secretary in Charge

Minoru Asada

Dept. of Adaptive Machine Systems,

Graduate School of Engineering,

Osaka University

2-1 Yamada-oka, Suita,

Osaka 565-0871, JAPAN /

JST ERATO Asada Synergistic

Intelligence Project

Secretary in Charge

Noriaki Mitsunaga

Dept. of Intelligent Robotics and

Communication Laboratories,

Advanced Telecommunications

Research Institute International

2-2-2 Hikaridai, Seika-cho, Souraku,

Kyoto 619-0288, JAPAN

Secretary

Kazuhiro Nakadai

Honda Research Institute Japan

8-1 Honcho, Wako,

Saitama, 351-0114, JAPAN /

Dept. of Math. and Comp. Sci.,

Graduate School of Engineering,

Tokyo Institute of Technology