Behaviour Learning System for Robot Soccer Using Neural Network

Moeko Tominaga (tominaga.moeko382@mail.kyutech.jp)

Kyushu Institute of Technology Graduate School of Life Science and Systems Engineering: Kyushu Kogyo Daigaku Daigakuin Seimeitai Kogaku Kenkyuka https://orcid.org/0000-0002-0500-4081

Yasunori Takemura

Nishinippon Institute of Technology: Nishinippon Kogyo Daigaku

Kazuo Ishii

Kyushu Institute of Technology Graduate School of Life Science and Systems Engineering: Kyushu Kogyo Daigaku Daigakuin Seimeitai Kogaku Kenkyuka

---

Research Article

Keywords: Autonomous behaviour selection, Neural network, Human-robot coordination

Posted Date: January 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-140970/v1

RESEARCH

Behaviour Learning System for Robot Soccer Using Neural Network

Moeko Tominaga[1*], Yasunori Takemura[2] and Kazuo Ishii[1]

[*]Correspondence: tominaga.moeko382@mail.kyutech.jp
[1]Graduate School of Life and Science Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, 808-0196 Kitakyushu, Japan
Full list of author information is available at the end of the article

Abstract

Technological developments have raised the promise of a human–robot symbiotic society. A soccer game has characteristics similar to those expected in such a society. Soccer is a multiagent game in which the strategy employed depends on each agent ' s position and actions. This study discusses the results of the development of a learning system that uses a self-organising map to select behaviours depending on the situation. This system can reproduce the action selection algorithm of all players in a certain team, and the robot can instantly select the next cooperative action from the information obtained during the game. In this manner, common sense rules can be shared to learn an action selection algorithm for a set of both human and robot agents as opposed to robots alone.

Keywords: Autonomous behaviour selection; Neural network; Human–robot coordination

Background

Robots are variously expected to realise a sustainable society, compensate for the worker shortage due to the rapidly aging society with low birth rate, and contribute to the reconstruction of industrial infrastructure. To realise a safe and secure robot society, research and development efforts should be directed toward satisfying social expectations and obtaining meaningful results. In particular, it is essential to discuss how to coexist with robots and thereby realise a symbiotic society. In a symbiotic society, humans and robots should ideally be able to interact with and understand each other, that is, they should be able to interact with their own behaviours as well as with all other autonomous agents that sense the environment and act according to autonomous policies.

In this light, the present study aims to develop an appropriate algorithm for intelligent robots that share an environment with humans. We use a soccer field as a test bed. Because soccer involves strategy, cooperation among players, unpredictable movements, and a common goal, it has many similarities to the behavioural algorithms of intelligent robots as described above, making it an ideal testbed for development.

We analysed the strategic behaviour of human soccer and compared it with that of robot soccer, focused on the differences in the cooperative behaviour of soccer between humans and robots, and analysed it using self-organizing maps (SOMs).[1] As a result, we found that in human soccer, the positioning of players was affected by their ranking in terms of distance from the ball within the team. Both experts who played regularly with the same team members as well as amateurs who formed

a team for the first time showed this behaviour. This suggests that individuals position themselves as a team according to their distance from the ball.
In the comparison between human and robot soccer matches, the positions of all players (except for the goalkeeper) and the ball were used as inputs, and the scenes were analysed up to the shooting action. As a result, we found that both human and robotic players played different team formations when they performed offensive and defensive actions. In addition, the human game was more finely divided into different situations than the robot game, with a distinction made between the time when the team was on defense from start to finish and the time when the team formation changed from defense to offense. This suggests that strategic behavioural choices are made in both human and robot games; however, human games show transitions between defensive and offensive formations.
Thus, in soccer matches played by multiple agents as a team, agents always position themselves for cooperation, and the positioning of surrounding players (including friend and foe) and the team formation are related to the action selection of a certain player of a certain team.
In this study, we develop a behaviour learning system for soccer robots based on these findings. This system learns the behaviour of all players (except the goalkeeper) in a team. The evaluation is based on the reproducibility of the learned behaviour selector. This system can reproduce the action selection algorithm of all players in a team, and the robot can instantly select the next cooperative action based on the information obtained during the game. For example, if the robot learns the match data of a team consisting of elementary school children, it can select the next action based on the behavioural criteria of this team. If it learns the match data of a European club team, it can strategically play the match with a formation specific to that team. For learning, we use a tensor self-organizing map (TSOM)[2]. In this way, if the action selection algorithm of a group of agents can be learned, "common sense rules'' can be shared with a team of robots as well as with a group of heterogeneous agents (e.g. humans and robots). This will play an important role in a future human-robot symbiotic society.
The remainder of this paper is organized as follows. First, we present an overview of recently studied learning algorithms for constructing a learning system. Next, we describe the algorithms and experimental methods of the SOM. Then, we describe the experiments, and finally, we summarise the obtained results.

Coordinated Action

Cooperative behaviour is an important aspect when different autonomous agents communicate with each other while executing a common task. Executing a task with a single agent is often inefficient. Therefore, researchers have recently been studying multiagent systems (MAS) to solve difficult problems. In a MAS, multiple autonomous agents try to achieve a common goal through cooperative behaviour. It interacts with both the environment and other agents by making real-time decisions based on sensor data[3][4]. RoboCup is a global robotics competition that focuses on multiagent coordination using learning methods such as reinforcement learning and neural networks. It aims to create a robot team that can beat the human soccer world cup champion team by 2050[5]. Sandholm and Crites[6] have shown that

reinforcement learning is an optimal method for addressing the iterative prisoner's dilemma provided that sufficient measurement and behavioural data are available. Furthermore, Arai[7] compared Q-learning and benefit-sharing methods for a tracking problem in a MAS in which the environment is modelled as a grid and showed that cooperative behaviour emerges clearly during benefit-sharing. However, these studies did not consider robots operating in real environments.

Robotic learning

Among various learning algorithms, unsupervised learning is promising for MAS systems. The advantage of unsupervised learning is that it does not require any previous knowledge of the robot itself, and the robot does not require any previous information about the environment. However, because learning requires considerable testing and experience, designers must clearly define the important parameters and variables to be considered during the match to define the state space and action space and thereby reduce the training cost and time. The design of the state space depends on the capabilities and action space of the robot as well as on what the robot can do in the action space. The two spaces are thus interconnected. To design the state space efficiently, Asada[8] proposed a method of first dividing the state space into two main states and then recursively dividing it into many layers while increasing the number of states. However, it was shown that some bias problems still affect the behaviour of the robot and may even cause some wrong behaviours.

Human–robot coordination

In a human–robot symbiotic system, robots must be able to understand and interpret human behaviours and to act accordingly. However, most previous studies have focused on only the robot behaviour or on behaviour based on a specific number of precisely defined interactions between humans and robots. In addition, in MASs, humans and robots usually communicate through an interface like voice or gesture commands. However, this may be insufficient in a very complex system. Therefore, the next step is to make the robot understand some situations and adapt its behaviour predictably in a manner similar to humans. This study aims to enable robots to show the same cooperative behaviour as humans. The elements of the input vector to the neural network are chosen so as to achieve the smallest possible gap between human and robot behaviours. Toward this end, human behaviour must first be understood, and then, a robot that can understand and mimic this behaviour must be developed. A previous study[9] investigated human and robot behaviours in futsal games because this, too, is a dynamic environment with some constraints and the need for real-time planning. A potential general human–robot symbiotic system of the future will have similar characteristics. The algorithm used analysed player positions using a SOM according to specific events in the game, such as score, corner kicks, penalty kicks, and team with the ball. As a result, the scene could be evaluated for both human and robot games, and increasing the number of input elements to be learned enabled a robot to develop autonomous behaviours by creating a SOM that shows a mapping relationship between the environment and the behaviour. This study develops a behaviour learning system for a soccer robot that learns matches between robots using a TSOM to find the mapping relationship.

Methods

Tensor self-organization map

An action learning system and an action selection system are developed using a TSOM. TSOM is an extension of SOM to tensor data. The TSOM algorithm involves three main steps: winner determination, nearest neighbour calculation, and model update (see Fig. 1).[10]

1 Winner decision

Eqs.1 and 2 are used to determine the best matching unit (BMU) for each mode, where k_i^* and l_i^* are the BMU coordinates for modes 1 and 2, respectively.

$$k_i^* = \arg\min_k \sum_{l=1}^{L} \sum_{d=1}^{D} (u_{ild} - y_{kld})^2 \tag{1}$$

$$l_j^* = \arg\min_l \sum_{k=1}^{K} \sum_{d=1}^{D} (v_{kjd} - y_{kld})^2 \tag{2}$$

2 Neighbourhood calculation

Eqs.3 and 4 are used to find the neighbourhood; α_{ki} and β_{lj} are the changes in the BMU neighbourhood unit for modes 1 and 2, respectively.

$$\alpha_{ki} = \exp(-\frac{1}{2\sigma_1^2} \|\zeta_{k_i^*}^{(1)} - \zeta_k^{(1)}\|^2) \tag{3}$$

$$\beta_{lj} = \exp(-\frac{1}{2\sigma_2^2} \|\zeta_{l_j^*}^{(2)} - \zeta_l^{(2)}\|^2) \tag{4}$$

3 Updating the model

Update both the first and second models using Eqs. 5–8.

- Update secondary model

$$y_{kld} = \frac{1}{g_k g_l} \sum_{i=1}^{I} \sum_{j=1}^{J} \alpha_{ki} \beta_{lj} x_{ijd} \tag{5}$$

$$g_k = \sum_{i=1}^{I} \alpha_{ki} \tag{6}$$

$$g_l = \sum_{j=1}^{J} \beta_{lj} \tag{7}$$

- Update primary model

$$u_{ild} = \frac{1}{g_l} \sum_{j=1}^{J} \beta_{lj} x_{ijd} \tag{8}$$

$$v_{kjd} = \frac{1}{g_k} \sum_{i=1}^{I} \alpha_{ki} x_{ijd} \tag{9}$$

These three steps are repeated while the neighbourhood radius σ is monotonically decreased according to the time constant τ.

Method of the experiment

In experiments, the simulation function used by "Hibikino-Musashi,'' the winning team of the RoboCup Medium League in Japan, is used to run the actual game program, and the action selection algorithm of robot soccer is learned based on the logged data. In the Medium League, only the sensors equipped on the robot are used, and no outside information is used. Further, the fold size is 14 m × 22 m, which is the largest in the RoboCup league. Currently, this is the only robot team league that can play soccer with humans. Because the actual game program is run in the simulation, there is no discrepancy between the robot's perception of the environment and its choice of action, and only the reproducibility can be evaluated. Open Dynamic Engine (ODE), a three-dimensional dynamics simulator developed for performing perform physical simulations, is used in this study. It is operated at 70 Hz, and log data are acquired at 15 Hz.

Learning requirement

The input data is a tensor, and the player, scene, and next action are extracted from the logged data. Table 1 shows the input vector x_{ijd} (see Figs. 1 and 2).
Here, $player\alpha$, $player\beta$, $player\gamma$, and $player\delta$ indicate the players closest to the ball in order, because in the current program of the Hibikino-Musashi team, the roles of the robots (FW, MF, and DF) are assigned according to the proximity of the ball to the robot at that moment. A series of 10 log data from kickoff to a shot were used as input data. "haveball'' is a binary value indicating whether the robot is holding the ball or not. "kick'' means that the robot ejects the ball, and its value increases with the amount of force used to eject the ball. All these input data were normalized for each writing element.

Table 2 shows the parameters used for learning using the TSOM. Fig. 3 shows that the number of training times I; maximum and minimum neighbourhood radius $\sigma_{max} and \sigma_{min}$, respectively; and time constant τ did not cause any problems in the convergence of the training.

Results and discussion

Figs. 4 and 5 show component plane views of the two feature maps obtained from the developed action learning system, and Fig. 6 shows the match situation on the simulator shown by the BMU at that moment.
This feature map shows a gradient of the magnitude of the *haveball* values of $player\alpha$. Red and blue values are high and low, respectively. In other words, this map shows that the game situation where $player\alpha$ is holding the ball is located in the lower right corner of the map.
Next, Fig. 5 shows the gradation of input element values when the data level of

$player\alpha$ is 10. In this case, A is holding the ball in the middle of the field, as shown in Fig. 6. At this time, the BMU of $player\alpha$'s *haveball* is red and that of *kick* is yellow, indicating that $player\alpha$ is holding the ball and is trying to output a weak ball ejection force as its next action. In other words, $player\alpha$ is about to pass the ball to a teammate. The learning results obtained using one TSOM from two feature maps are thus accurate.

Conclusions

This study develops a behaviour learning system for a soccer robot. The data obtained from this system are stored in the robot, and when the BMU is selected from the data obtained during the match, the robot outputs a plausible behaviour as a result of learning. In other words, the behaviour learned using this behaviour learning system can be selected. This result indicates that team actions can be performed with multiple heterogeneous agents.

In the future, we will verify whether more complex behaviours can be reproduced and conduct experiments on actual robots that use the behaviour selection algorithm obtained using the learning system.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and material
The datasets analysed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Funding
Not applicable

Author's contributions
The first author designed and conducted the study, analysed and interpreted the data, and wrote the manuscript. The second author supported the idea and performed the programming. The third author supported the programming and conceptualised this study.

Acknowledgements
We would like to thank Mr.Enrico di Maria and Editage (www.editage.com) for English language editing.

Author details
[1]Graduate School of Life and Science Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, 808-0196 Kitakyushu, Japan. [2]Department of Engineering, Nishinippon Institute of Technology, 1-11 Aratsu, Kanda-town, Miyako-gun, 800-0397 Fukuoka, Japan.

References

1. M. Tominaga, Y. Takemura, K. Ishii, "Analysis of Cooperative Behavior in Futsal for Human and Robot Multi Agent System '' , The Proceedings of JSME Annual Conference on Robotics and Mechatronics (Robomec), 2P2-L09, 2018.
2. T. Kohonen, "Self-Organizing Maps '' , Springer Japan, 2001.
3. P. Stone, M. Veloso: "Multiagent Systems: A Survey from a Machine Learning Perspective '' , Autonomous Robots, 8(3), pp. 345-383, 2000.
4. K. Ueda: "Souhatsu to Multi Agent System [Emergence and Multi Agent System] '' , Baifukan, 2007.
5. S. Behnke, R. Sheh, S. Sariel, D. D. Lee (Eds.), "RoboCup-2016: Robot World Cup XX '' , Springer-Verlag, 2017.
6. T. W. Sandholm, R. H. Crites, "On Multiagent Q-learning in a Semi-Competitive Domain '' , Workshop Notes of Adaptation and Learning in Multiagent Systems Workshop, IJCAI-95, 1995.
7. S. Arai, K. Miyazaki, S. Kobayashi, "Methodology in Multi-Agent Reinforcement Learning: Approaches by Q-Learning and Profit Sharing '' , Journal of Japanese Society for Artificial Intelligence, 13(4), pp. 609-618, 1998.
8. Y. Takahashi, M. Asada, "Behavior Acquisition by Multi-Layered Reinforcement Learning '' , Journal of the Robotics Society of Japan, 18(7), pp. 1040-1046, 2000.

9. M. Tominaga, Y.Takemura, K.Ishii, "Behavior Analysis of Multi-Agent Game for Human-robot Cooperation '' , 18th International Conference on Control, Automation and Systems, Korea, 2018.
10. T. Iwasaki, T. Furukawa, "Relational data visualization by Tensor SOM '' ,31st Fuzzy System Symposium, pp. 444-447, 2015.
11. A. Ultsch, H. P. Siemon, "Kohonen ' s Self Organizing Feature Maps for Exploratory Data Analysis '' , INNC'90, pp. 305-308, 1990.

Figures

Figure 1 Data image of TSOM

Figure 2 Environment for determining game situation

Figure 3 Neighbourhood radius and unit variation in epoch

Figure 4 Component plane display of BMU distribution and $player\alpha$'s $haveball$ values for 520 data

Figure 5 Component plane display of BMU distribution and $player\alpha$'s data level of 10 for 55 parameters

Figure 6 Game situation in simulator

Tables

Table 1 Parameters of TSOM

Number of input data	I (520 dimensions)	
State and Next action	J (55 dimensions)	*TP*, *B*, *NA1*, *NA2*, *NA3*, *NO1*, *NO2*, *NO3*, *NO4*, $Target_x$, $Target_y$, $Target_\theta$, $kick$
Role of players	D (4 dimensions)	$player\alpha$, $player\beta$, $player\gamma$, $player\delta$

Table 2 Parameters of TSOM

Map size (mode 1)	K	10×10
Map size (mode 2)	L	10×10
Maximum neighbourhood radius (modes 1 and 2)	σ_1	2.0
Minimum neighbourhood radius (modes 1 and 2)	σ_1	0.2
Time constant	τ	100
Learning time	$epoch$	200

Additional Files

Figures

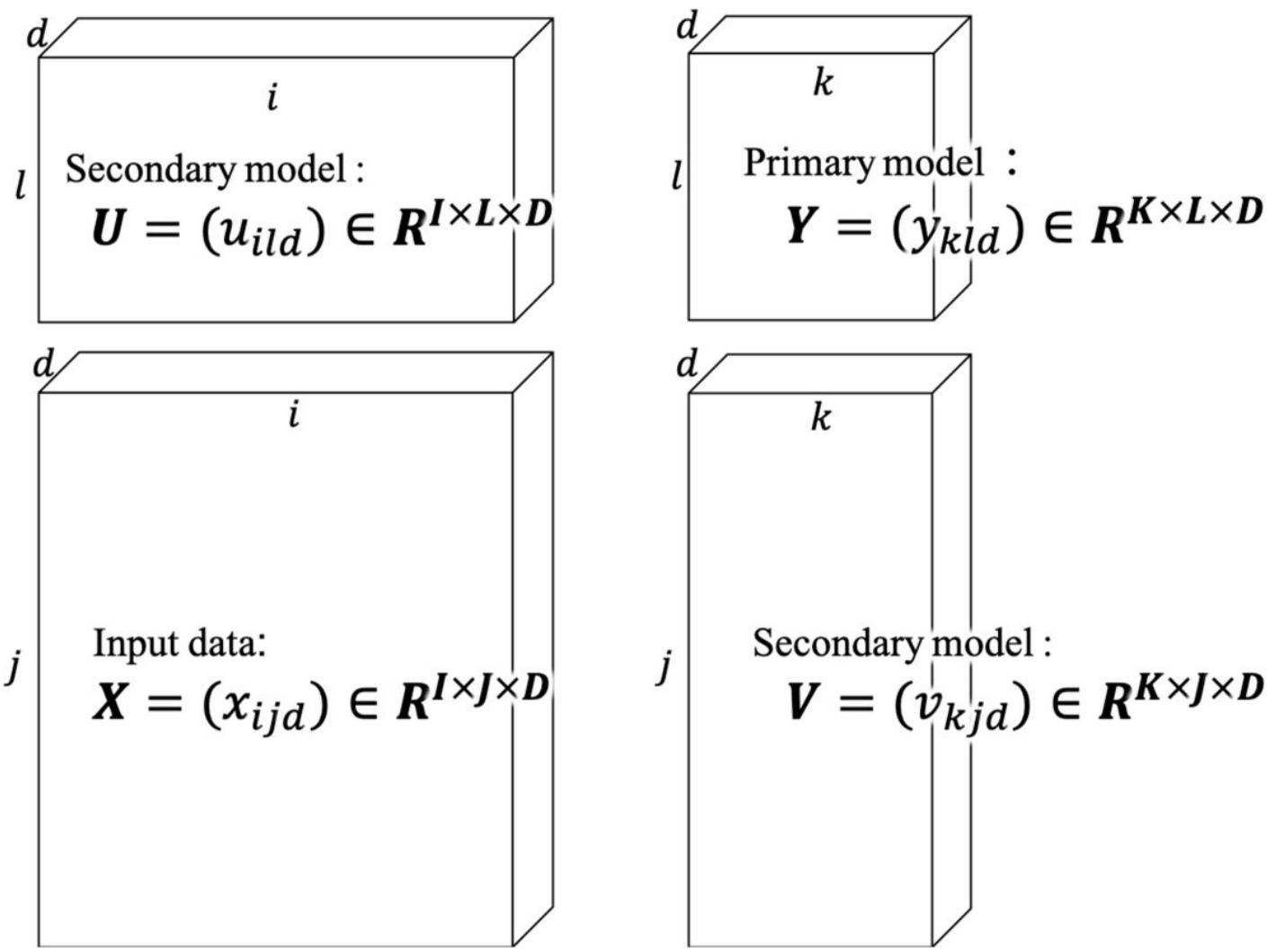

Figure 1

Data image of TSOM

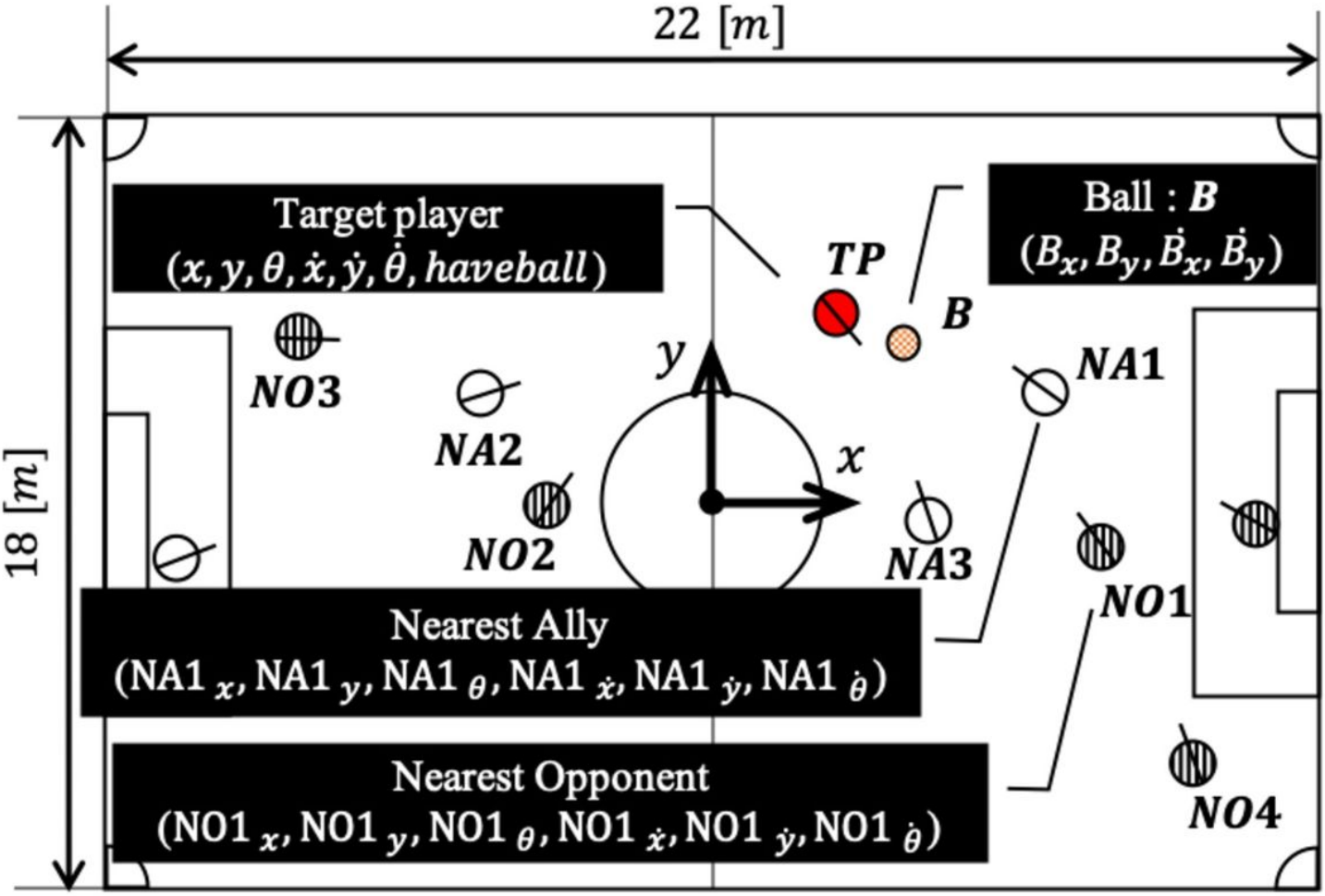

Figure 2

Environment for determining game situation

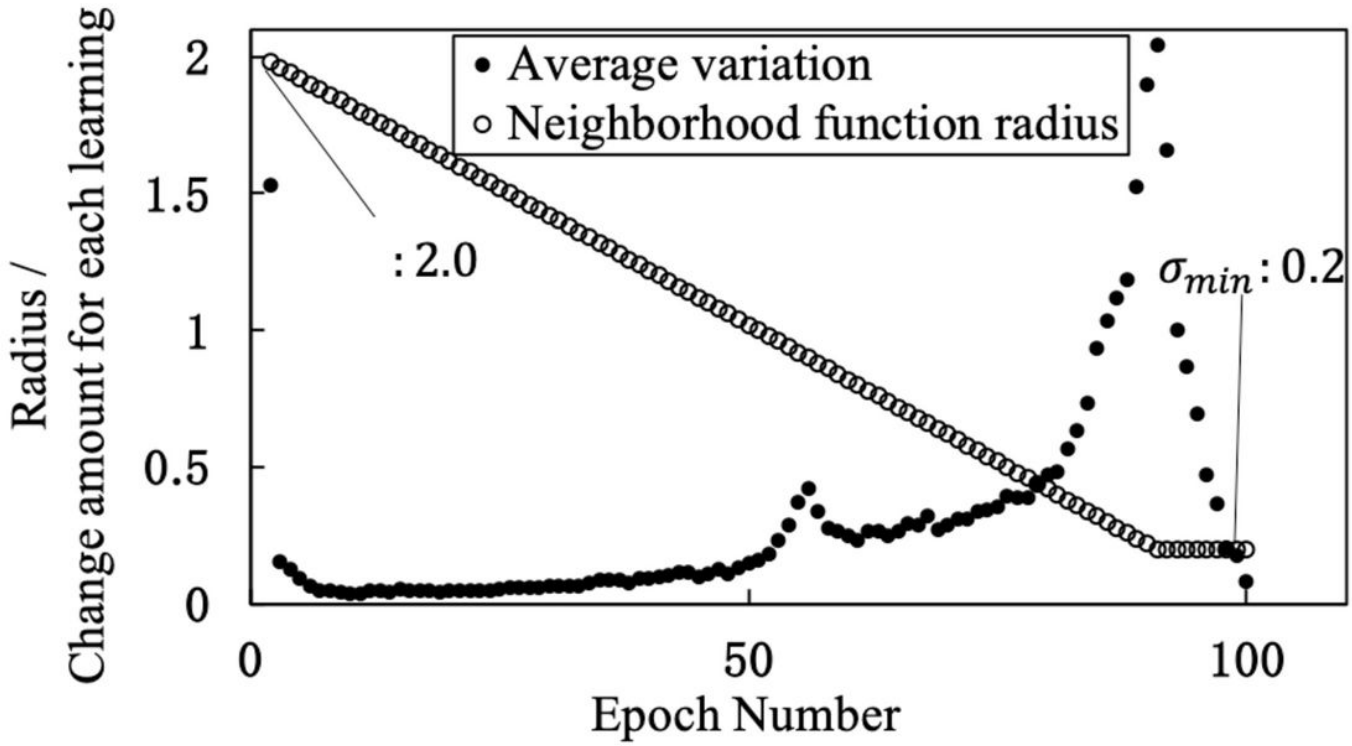

Figure 3

Neighbourhood radius and unit variation in epoch

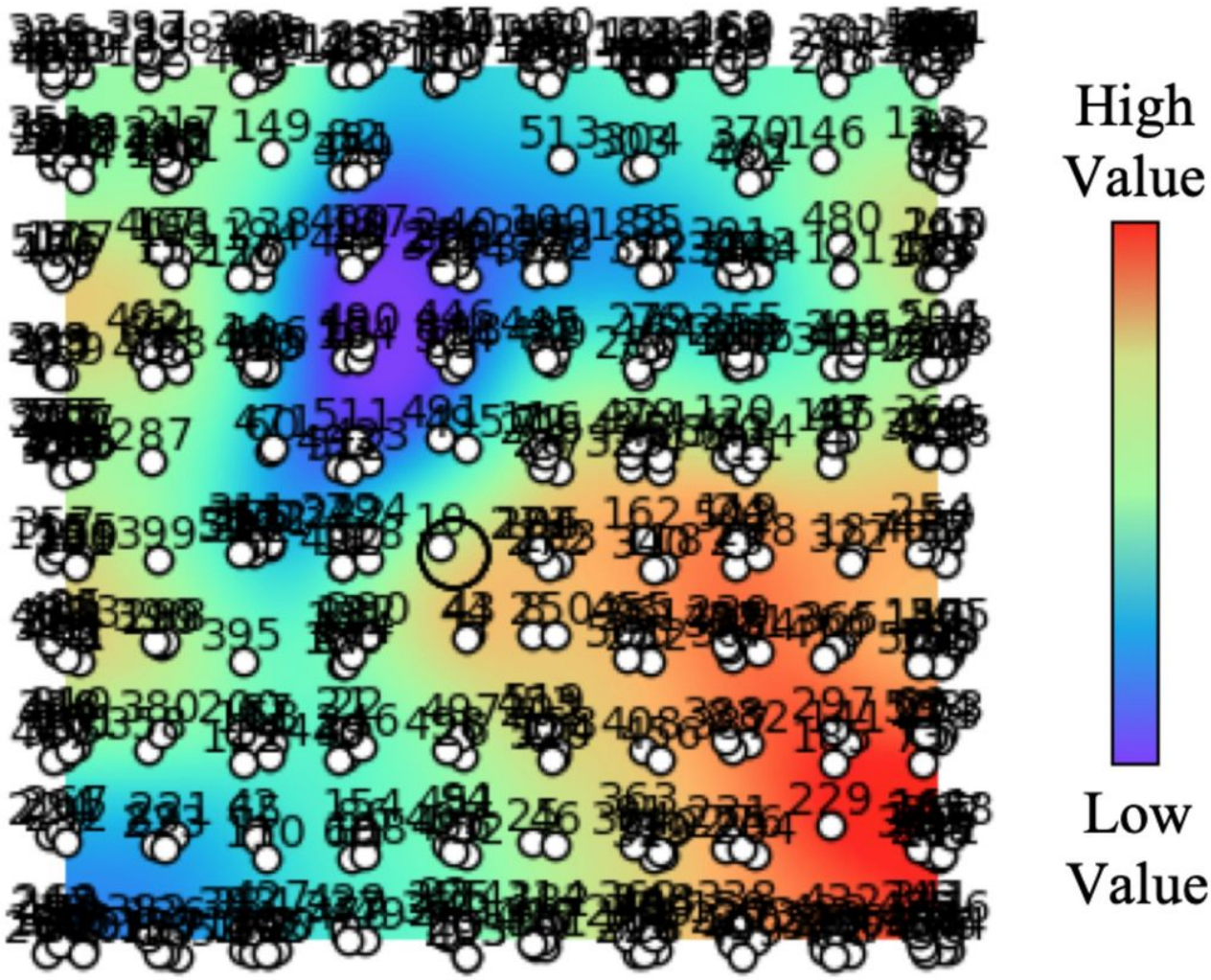

Figure 4

Component plane display of BMU distribution and playerα's haveball values for 520 data

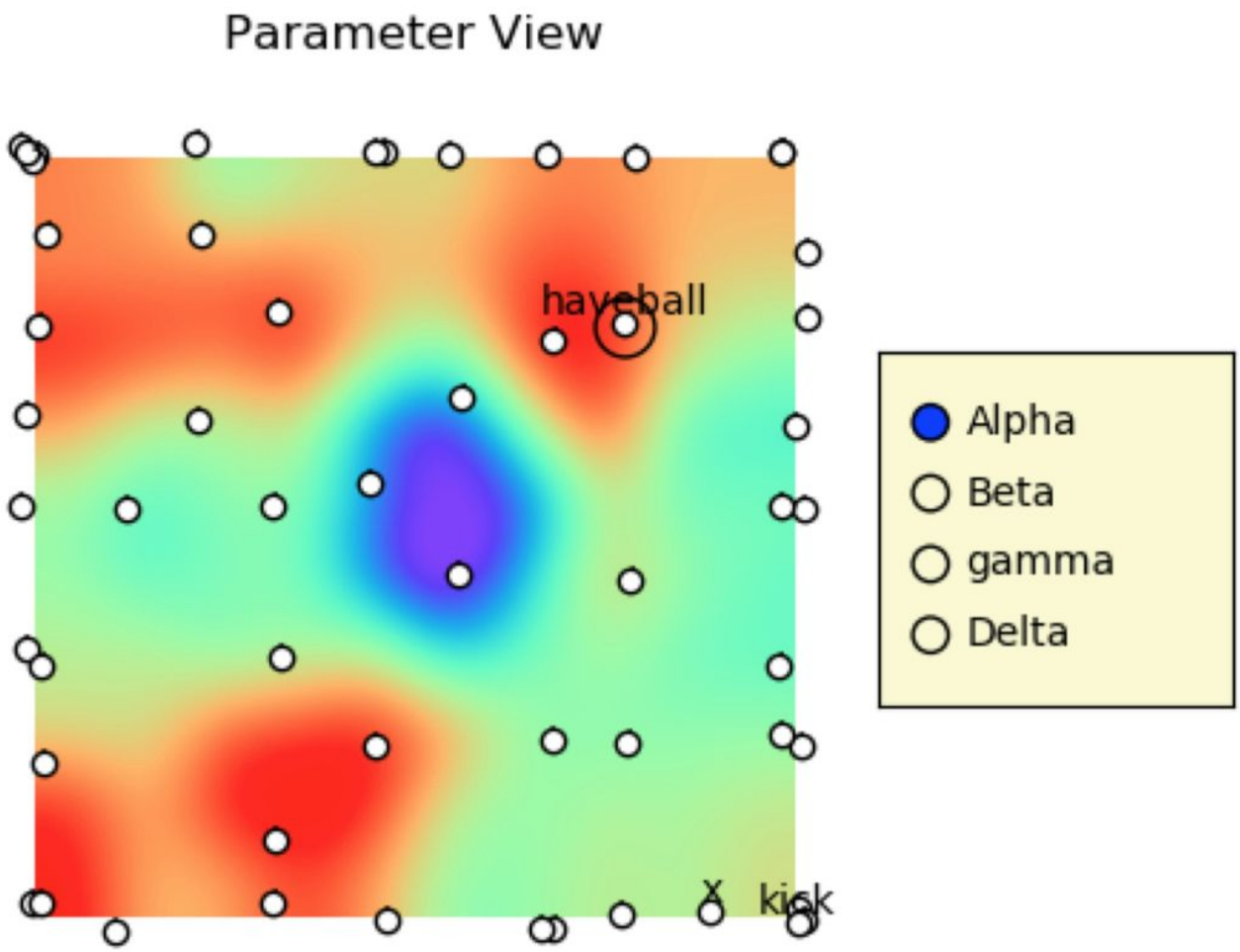

Figure 5

Component plane display of BMU distribution and playerα's data level of 10 for 55 parameters

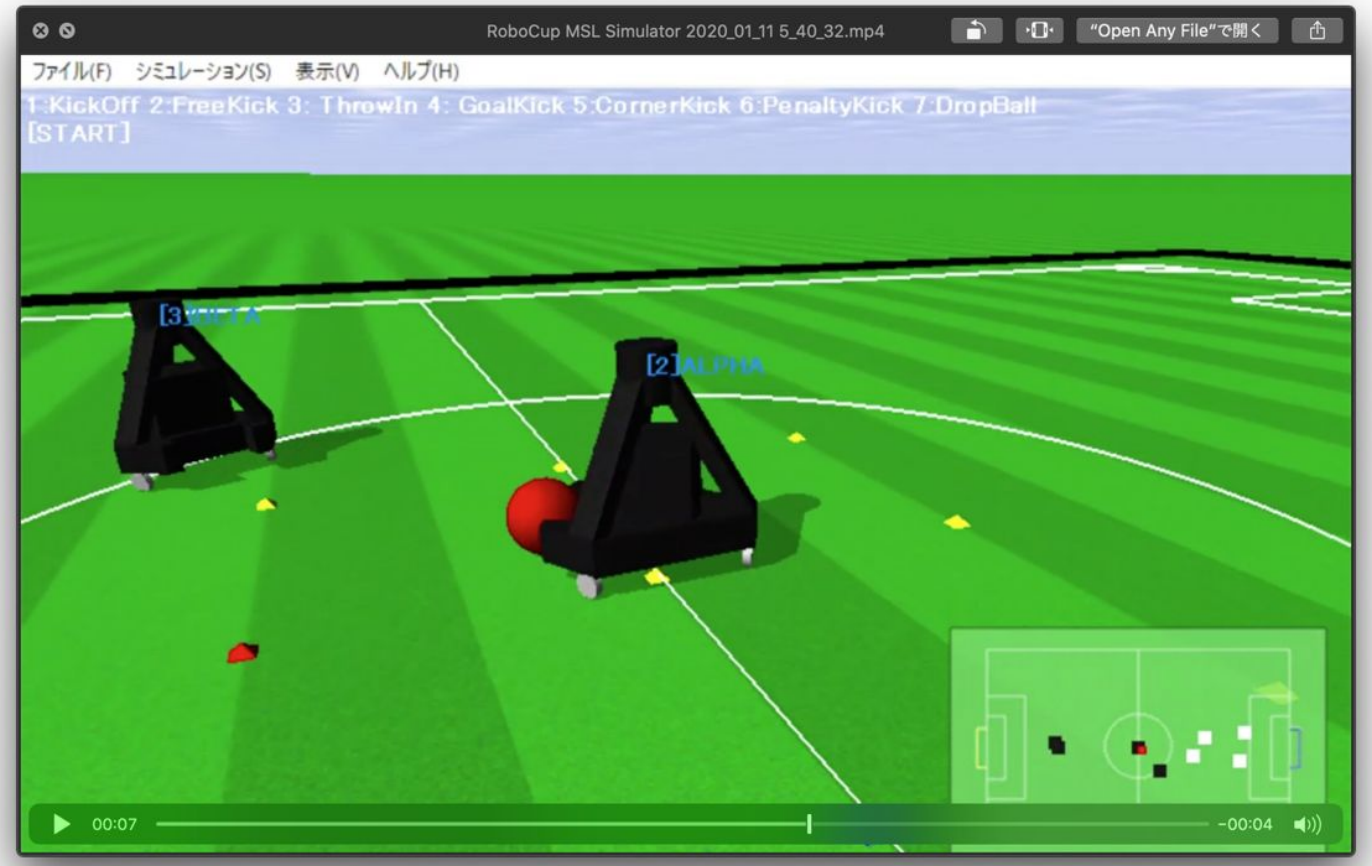

Figure 6

Game situation in simulator