ICIC Express Letters
Volume **13**, Number **4**, April **2019**

ICIC International ©2019 ISSN 1881-803X
pp. **303–310**

A STUDY ON THE ANALYSIS OF SOCCER GAMES USING DISTRIBUTED REPRESENTATION OF ACTIONS AND PLAYERS

Jiarun Zhong[1], Tomoharu Nakashima[1] and Hidehisa Akiyama[2]

[1]Graduate School of Humanities and Sustainable System Sciences
Osaka Prefecture University
Gakuencho 1-1, Naka-ku, Sakai, Osaka 599-8531, Japan
mbb04011@edu.osakafu-u.ac.jp; tomoharu.nakashima@kis.osakafu-u.ac.jp

[2]Department of Electronics Engineering and Computer Science
Faculty of Engineering
Fukuoka University
8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan
akym@fukuoka-u.ac.jp

Received September 2018; accepted December 2018

Abstract. *In this paper we propose a new method for analyzing team strategies from the action records during a game in RoboCup soccer simulation 2D league. In the soccer simulation 2D games, log files are mainly used to analyze the strategies of teams. Being able to design appropriate strategies of teams, we need to identify the performance of players and relationship between players in the teams. For this purpose, firstly, we use a natural language processing method which is called Word2vec. Word2vec translates the action and players into real vectors from log files. The real vectors are then used to perform clustering of players. As different teams may have different strategies, according to the clustering results of players, it is possible to grasp the similarity of teams, players, and then choose an appropriate strategy to play against the opponent team. With the proposed method, we also conduct some experiments with different teams and show effectiveness of the proposed method.*
Keywords: Sports data analysis, RoboCup soccer simulation 2D, Word2vec, Clustering

1. **Introduction.** In RoboCup soccer simulation 2D games, one of the essential parts in developing a team is to analyze logged information of games. We can set appropriate strategies to win a competition if we know how the players in the team move throughout games.

However, many researches on the analysis of RoboCup soccer 2D games mainly focus on characteristic quantities, such as possession percentage of the ball, the number of passes, dribbles, and shoots. These cannot show any clear connection between players and/or between teams. This is because it is hard to really understand the performance of a player only by paying attention to these characteristic quantities. For this reason, according to the logged information we need a method which shows the information of players' performance during the games. This will well represent the strategy formation of teams and the relationship between players and teams, and then decide a suitable strategy to win the game.

Therefore, this paper proposes a method which identifies some relationships between any players in a team and can also indicate some similarities between teams. To achieve this, at the first step, we also need some characteristic quantities. In this paper we only focus on pass and dribble as the features. We extract passes and dribbles from log information and save into a corpus, and then use Word2vec [1-3] which can translate

DOI: 10.24507/icicel.13.04.303

actions and players into real vectors. Get the vectors used for the following clustering step. Being processed by clustering the players will be divided into several groups. Players in the same group have strong relationship during games. Based on the group distribution of players we calculate distance of each different teams. Finally, according to the team distance space, we can know the similarity of any teams.

This paper is organized as follows. Section 2 describes what the RoboCup is and also some more details on the RoboCup soccer simulation 2D league are explained. Section 3 introduces the proposed method in detail. In Section 4 we conduct some experiments to show the results of our proposed method. Section 5 is the conclusion of this paper.

2. **RoboCup.** RoboCup [4] is an annual international scientific robot competition in which teams of multiple robots compete against each other. RoboCup soccer simulation 2D league is one of the competition groups in RoboCup.

A RoboCup soccer simulation 2D game is simulated on computers that calculate the physics of the objects such as the ball and the players. The simulator, which is called a soccer server, generates two types of log files at the end of each game. The log files comprise information about the game, in particular about the current status of all players and the ball including their velocity and orientation at each cycle. The actions available for players are body-turn (change the players' body direction), neck-turn (change the players' view angle), dash (used to accelerate the player in direction of its body), kick and others. A soccer simulation game in the RoboCup simulation 2D league lasts for ten minutes in total except for over time. The simulation time is discrete, that is, a game consists of 6000 cycles where the length of each cycle is 100ms. In this paper we will concern the actions such as pass and dribble which involve primary actions such as kick and dash commands.

3. **Proposed Method.** The proposed method in this paper consists of four main components: (1) to extract actions from log files, (2) to obtain the vectors of actions and players based on Word2vec, (3) to perform clustering for actions and players, and (4) finally construct a distance matrix for teams. Figure 1 shows the overall structure of the proposed method.

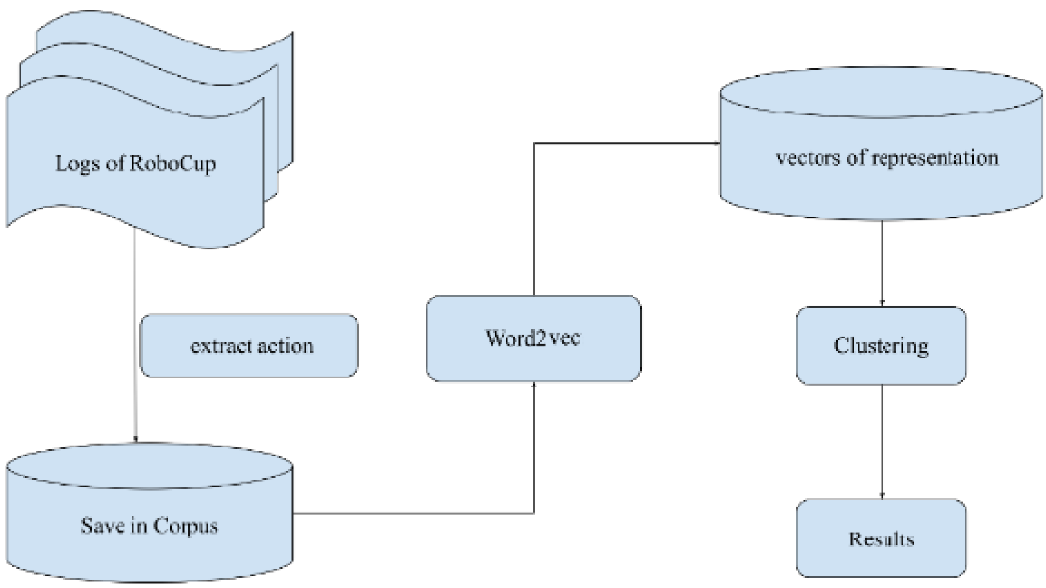

Figure 1. Framework of the proposed method

3.1. **Action extraction based on log files.** This is a key step to analyze game logs. Although the soccer simulator records games in log files, it is difficult to directly analyze them because the information in the log files is too raw for human being to understand. In order to analyze the games using a natural language processing method, a corpus should be generated. A corpus includes the series of actions of players during a game. In the proposed method, an action is considered to be a word in a sentence, and a sequence of actions can be considered as a sentence in the field of natural language processing. A series of actions is defined as a sequence of actions of a single team without any intervention by the other team. In this paper, we only consider pass and dribble actions. The other actions such as shoot, tackle, and intercept are ignored. In this paper a pass is defined as follows.

$$Pass : Kick_{(i)} - Kick_{(j)} \tag{1}$$

This means that a pass is defined as two subsequent kick commands by different players from the same team. A dribble is defined as follows.

$$Dribble : Kick_{(i)} - Dash_{(i)} - Kick_{(i)} \tag{2}$$

A dribble is defined as more than one succeeding combination of kick and dash commands by the same player. The indices $i, j = 1, \ldots, 11$ in (1) and (2) are the player's uniform number.

And we set a constant min_count in corpus to 15, which means if the player or action less than 15 times in a corpus it will be ignored. For example, there is no one pass between goalie and one of team members. This situation does not help us to analyze games. So, it should be ignored.

Figure 2 shows an example of extracted actions from a log file of a game. In this figure, different colors represent different players. To save the actions into corpus, we consider a sequence of actions as one line, which means that from the time the team gets the ball until the ball is intercepted by its opponent team, all the actions are treated in one line in the corpus. By this processing, we can obtain a corpus which includes the description of the team actions.

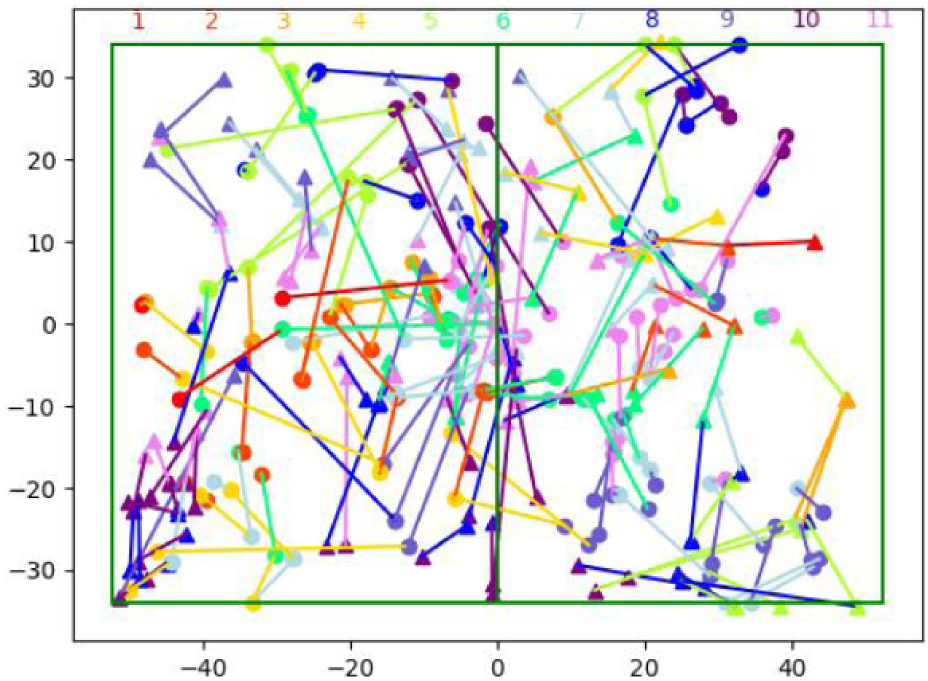

FIGURE 2. Example of extracted actions

3.2. **Learning representation by Word2vec.** By the procedure described in the last subsection, we obtain a corpus of actions. In order to convert actions to real vectors, we use Word2vec [1-3] using the generated corpus to train neural network.

Word2vec is a tool often used in the field of natural language processing. It has been applied to many research tasks such as sentiment analysis [7,10] and text classification [8]. Word2vec uses a set of one-to-one mapping model from a word to a real vector. In the proposed method, the actions and the number of players are extracted and included in the corpus as described in the last subsection. Otherwise, the actions or players cannot be converted to real vectors. In Word2vec, it is possible to discover the relation among words in the corpus. Thus, it can be applied to the analysis of simulated soccer games. The dimensionality of the real vector should be set large enough so that the distance between actions and players can be grasped easily by human being. In this step, not only actions but also players (in fact players' uniform numbers) are converted to their corresponding real vectors.

Figure 3 shows an example of converting an action and player to a real vector. Player 4 passes the ball to a player who is in front of him. Before processing by Word2vec, the action and player are recorded as strings in the corpus as in Figure 3. Then, Player 4 and the pass action are translated into vectors as in the right section of Figure 3.

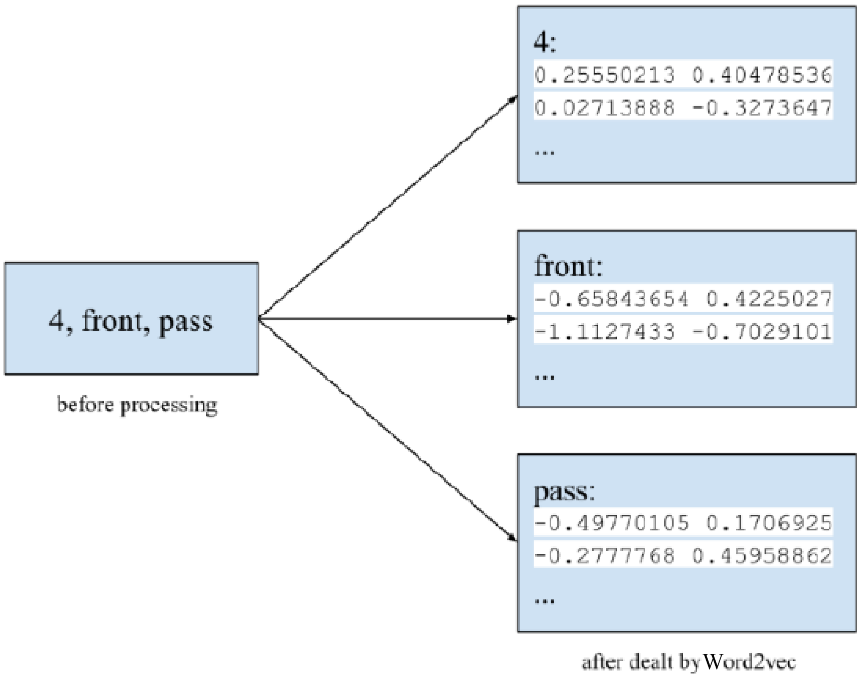

Figure 3. An example of converting an action to a vector

3.3. **Clustering.** As described above, we extract the actions of all players and convert them to real vectors in order to investigate the team performance and the relationship between actions and players. For the clustering, we pay attention only to players.

In the step of clustering, the number of clusters k is set to 5. In the previous section (Subsection 3.2), after processing by Word2vec, we obtained the vector of each player, through which we used k-means to apply the clustering to the vectors that are corresponding to the players. The clustering results divide the players into five groups. It is seen that the players in each group are highly correlated. Then we can calculate the distance [11] between each team in space based on the grouping results of players, so as to measure the similarity between teams by distance, indicating that the team is very similar. In the end, when we get to distance space, we do one more clustering, so we can separate the teams from the other team which is far away. Of course we can also know the similarity of teams from the result.

Figure 4 indicates the information what we can obtain from clustering. From the first result of clustering we can see that the players of the team are divided into different groups, which means that these players have strong connection in the same group. The second clustering based on distribution of players gets from the first step. Finally, it will be shown which teams are similar to each other.

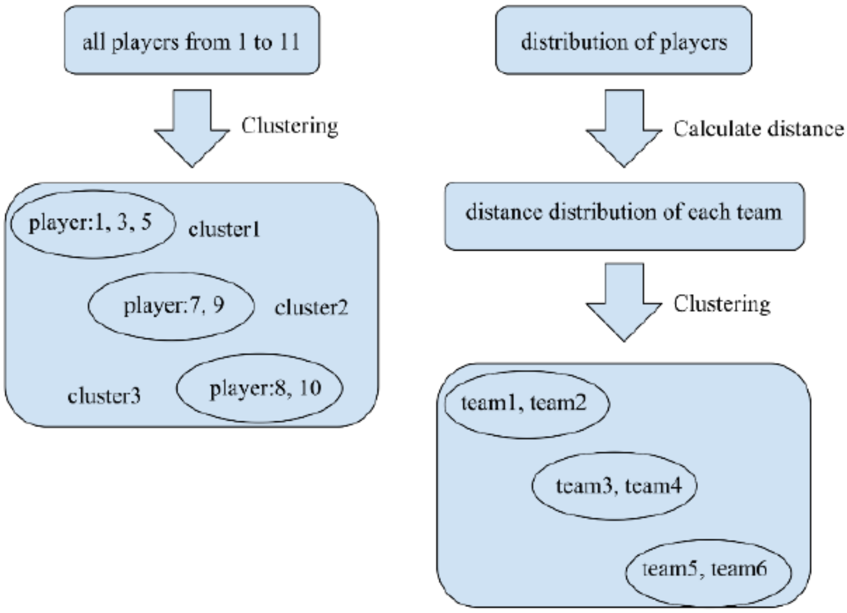

Figure 4. Two processes of clustering

4. **Experiment and Analysis.** In this section, we conduct a series of experiments to show the performance of our proposed method and examine the results if useful information can be discovered or not.

4.1. **Data preparation.** For evaluating our method, we need to prepare the data set that is necessary for the analysis. We use recorded log files from 2D soccer simulation matches among six teams. The main data set that we use throughout this paper are log files that were generated from 1,500 matches among the following teams: opuSCOM2018, Agent2D [12], Ri-one2017 [13], HillStone [14], HELIOS2017 [15] and Fifty-Storms [16]. Each team had 100 rounds of games against the other teams. A corpus was generated so that it only contains the actions and players from the same team. That is to say, one corpus corresponds to only one team and six corpuses were generated in our experiments. In this experiment, we will analyze each team's game separately, and then compare each result.

4.2. **Clustering of games.** We will do clustering twice in the experiment. First is clustering for the elements such as actions and players, and the other is for finding which team is similar to which team. By the first clustering based on the real vectors of players and actions, we construct a matrix that shows if any players or actions are in the same cluster (1) or not (0). The diagonal element (i.e., the identical player or action) is set to 0. Thus, this generates a clustering space which represents the relationship between the actions and players in the team. Table 1 shows the clustering results for Team opuSCOM2018.

As we can see from Table 1, Player 4 and Player 7 are in the same cluster because plenty of passes were done between these two players in whole game logs. In other words,

Table 1. Clustering result of elements (opuSCOM2018)

Num	1	2	3	4	5	6	7	8	9	10	11
1	0	1	1	0	1	0	0	0	0	0	0
2	1	0	1	0	1	0	0	0	0	0	0
3	1	1	0	0	1	0	0	0	0	0	0
4	0	0	0	0	0	0	1	0	0	0	0
5	1	1	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	1	0	0	0
7	0	0	0	1	0	0	0	0	0	0	0
8	0	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0

Table 2. Distance of teams without dribble

	opuSCOM	Ri-one	HillStone	Agent2D	HELIOS	Fifty-Storms
opuSCOM	0	4.69041576	5.099019514	2.449489743	6.0	3.16227766
Ri-one	4.69041576	0	4.898979486	6.164414003	5.291502622	5.656854249
HillStone	5.099019514	4.898979486	0	4.898979486	4.472135955	5.656854249
Agent2D	2.449489743	6.164414003	4.898979486	0	5.291502622	4.898979486
HELIOS	6.0	5.291502622	4.472135955	5.291502622	0	4.69041576
Fifty-Storms	3.16227766	5.656854249	5.656854249	4.898979486	4.69041576	0

Table 3. Clustering of similarity without dribble

	opuSCOM	Ri-one	HillStone	Agent2D	HELIOS	Fifty-Storms
opuSCOM		0	0	1	0	1
Ri-one	0		0	0	0	0
HillStone	0	0		0	1	0
Agent2D	1	0	0		0	1
HELIOS	0	0	1	0		0
Fifty-Storms	1	0	0	1	0	

Table 4. Distance of teams with dribble

	opuSCOM	Ri-one	HillStone	Agent2D	HELIOS	Fifty-Storms
opuSCOM	0	5.656854249	4.242640687	1.414213562	3.16227766	5.656854249
Ri-one	5.656854249	0	3.741657387	4.242640687	2.828427125	4.472135955
HillStone	4.242640687	3.741657387	0	4.242640687	4.472135955	4.242640687
Agent2D	1.414213562	4.242640687	4.242640687	0	4.242640687	4.0
HELIOS	3.16227766	2.828427125	4.472135955	4.242640687	0	5.830951895
Fifty-Storms	5.656854249	4.472135955	4.242640687	4.0	5.830951895	0

it becomes clear that these two players have a strong relationship in this team. After obtaining the clustering result for all the teams, the distances of each pair of teams are calculated. Thus, a distance matrix of the teams is generated. Clustering of teams according to distance matrix reveals the similarity of teams. Tables 2-5 show the results of the proposed method.

4.3. **Discussion.** The conversion of actions and players into real values depends on the training corpus. Thus, the number of game logs and the parameters in Word2vec has an impact on the generation of the real vectors. Even if the same parameters are set,

Table 5. Clustering of similarity with dribble

	opuSCOM	Ri-one	HillStone	Agent2D	HELIOS	Fifty-Storms
opuSCOM		0	0	1	0	0
Ri-one	0		0	0	1	0
HillStone	0	0		0	0	1
Agent2D	1	0	0		0	0
HELIOS	0	1	0	0		0
Fifty-Storms	0	0	1	0	0	

different formations of team strategy can also result in different clustering results due to the randomness in the initialization process of the neural networks in Word2vec.

In our experiments, the dimension of the vectors is set to 200, and the number of training iterations is set to 100.

5. **Conclusions.** In this paper, we proposed a new method to analyze simulation 2D soccer games. This paper presented the possibility of identifying the performance of players and the similarity of teams. A sequence of actions among players is the focus on this research. A single action without any player information is not enough to obtain useful information on the strategy of teams and even do not know how to improve the ability of team for future competition. This paper proposed a method that converts the binary format actions and players to translate into text corpus and then used Word2vec to convert it to real vectors. At the final phase use k-means to clustering based real vectors of players and actions to generate group of distribution. Use the distribution to calculate the distance of teams, which can find a team is similar to other teams.

However, we need more experiments to confirm the results whether any two teams are really similar in the future. The method of this paper is in the beginning of the research work. The next step is to further improve the certainty of clustering result based on players and use other teams. In the experiments of this paper, the number of players is fixed. If the teams have different types of players' number for different games, this will result in totally different consequence. Furthermore, the results of this study should be used in the actual analysis task of games. Certainly there is another future research, which is to extract more useful information or actions such as tackle shoot from log files.

REFERENCES

[1] T. Mikolov, G. Corrado, K. Chen and J. Dean, Efficient estimation of word representations in vector space, *Proc. of the International Conference on Learning Representations Workshop*, pp.1-12, 2013.

[2] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, Distributed representations of words and phrases and their compositionality, *Proc. of Advances in Neural Information Processing Systems*, pp.3111-3119, 2013.

[3] Q. Le and T. Mikolov, Distributed representation of sentences and documents, *Proc. of the International Conference on Machine Learning*, pp.1188-1196, 2014.

[4] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa and H. Matsubara, RoboCup: A challenge problem for AI, *AI Magazine*, vol.18, no.1, pp.73-85, 1997.

[5] X. Lin, H. Li, Y. Zhang, L. Gao, L. Zhao and M. Deng, A probabilisitc embedding clustering method for urban structure detection, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W7, pp.1263-1268, 2017.

[6] B. M. Faria, L. P. Reis, N. Lau and G. Castillo, Machine learning algorithms applied to the classification of robotic soccer formations and opponent teams, *Proc. of the IEEE Cybernetics and Intelligent Systems*, pp.344-349, 2010.

[7] D. Mayank, K. Padmanabhan and K. Pal, Multi-sentiment modeling with scalable systematic labeled data generation via Word2vec clustering, *Proc. of the 16th IEEE International Conference on Data Mining Workshops*, pp.952-959, 2016.

[8] J. Lilleberg, Y. Zhu and Y. Zhang, Support vector machines and Word2vec for text classification with semantic features, *Proc. of the 14th IEEE International Conference on Cognitive Informatics & Cognitive Computing*, pp.136-140, 2015.
[9] O. Michael, O. Obst, F. Schmidsberger and F. Stolzenburg, Analysing soccer game with clustering and conceptors, *Proc. of RoboCup Symposium 2017*, Nagoya, Japan, pp.1-12, 2017.
[10] E. M. Alshari, A. Azman, S. Doraisamy, N. Mustapha and M. Alkeshr, Improvement of sentiment analysis based on clustering of Word2vec features, *Proc. of the 28th International Workshop on Database and Expert Systems Application*, pp.123-126, 2017.
[11] T. Nakashima, M. Ogura, R. Yoshida, T. Nishijima, M. Ibuki, M. Miyabe and T. Kawamata, Construction of a vector space using item description –Hit product explorer of sweets and snacks–, *Proc. of the 6th Asian Conference on Information Systems*, Phnom Penh, Cambodia, pp.157-160, 2017.
[12] H. Akiyama and T. Nakashima, HELIOS_base: An open source package for the RoboCup soccer 2D simulation, *Proc. of RoboCup Symposium 2013*, pp.1-8, 2013.
[13] M. Mizumoto, T. Fuzimitsu, T. Ebara, S. Yamamoto, H. Asai, A. Ishida, S. Inoue, H. Oe, Y. Kawakami, T. Kitamura, N. Kitamura, S. Kono, K. Kobayashi, S. Takeda, T. Naito and Y. Hosomi, RoboCup 2017 – 2D soccer simulation league team description Ri-one, *Proc. of RoboCup Symposium and Competition*, Nagoya, Japan, 2017.
[14] T. Kiura, T. Omori and N. Watanabe, Team HillStone2017 in the 2D simulation league team description paper, *Proc. of RoboCup Symposium and Competition*, Nagoya, Japan, 2017.
[15] H. Akiyam, T. Nakashima, S. Tanaka and T. Fukushima, HELIOS2017: Team description paper, *Proc. of RoboCup Symposium and Competition*, Nagoya, Japan, 2017.
[16] H. Igarashi, J. Yamagishi and M. Irikura, Fifty-Storms2017: Team description paper, *Proc. of RoboCup Symposium and Competition*, Nagoya, Japan, 2017.