Cooperation Level Estimation of Pair Work Using Top-view Image

Katsuya Sakaguchi[1] and Kazutaka Shimada[1]

Kyushu Institute of Technology, 680-4 Iizuka Fukuoka, Japan
{k_sakaguchi, shimada}@pluto.ai.kyutech.ac.jp

Abstract. To understand an interaction among persons is one of the most important tasks in artificial intelligence. In this paper, we propose a method for estimating a cooperation level in pair work. The task is a cooperation work that take place in front of a whiteboard by two persons. The goal of our study is to provide the cooperation level that is estimated by features extracted from images for teachers. The result of this study is useful for education support systems and problem based learning. We extract the standing location, operation ratio and head direction of each person from an overhead camera. We apply the features to two machine learning approaches: AdaBooost and multiple linear regression. We obtained 77.5% as the accuracy by the AdaBoost and 0.649 as the adjusted R^2 by the regression.

Keywords: Interaction analysis, Cooperation Level, Pair Work, Top-view Image

1 Introduction

To understand an interaction among persons is one of the most important research tasks in artificial intelligence. We have proposed methods for understanding interactions in a conversation with spontaneous utterances [12, 14, 19]. In these studies, we focus on linguistic, phonetic and prosodic features. Utilizing information extracted from images is, however, necessary for understanding an interaction. Image data contains much information that linguistic information does not contain. Vargas [16] has reported that posture and gaze information are effective elements for estimating speaker's mind as the regulator that is actions such as a nod and a prompt of the next utterance. Mahmoud et al. [8] have reported an analysis of hand-over-face gestures for automatic inference of cognitive mental states. Kumano et al. [7] have analyzed how empathy and antipathy, aroused between people while interacting in face-to-face conversation, are perceived by external observers. In this paper, we also analyze an interaction with two persons by using features captured from images.

Recently, faculty development, which is to improve skills and knowledge about teaching ability, has been more important. For the purpose, Yamane et al. [18] have proposed a method to detect an interaction between a lecturer and learners. Furthermore, problem-based learning (PBL) is the more recent and

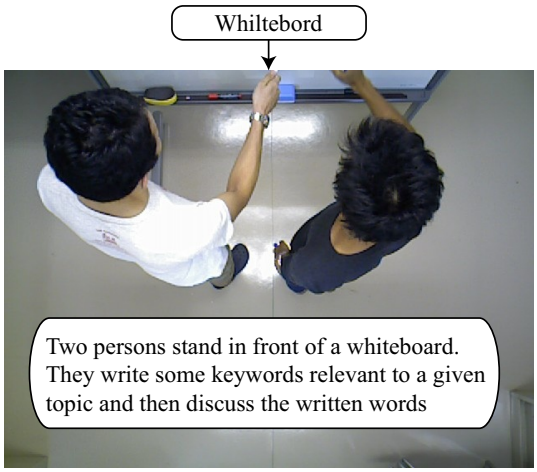

Fig. 1. An image of pair work from top-view.

highly regarded. In PBL, students work in small collaborative groups and learn what they need to know in order to solve a problem [3]. The target of our research is the PBL environment.

In this paper, we propose a method for estimating a cooperation level in pair work. The cooperation level in this study is that "To what degree did a participant work in cooperation with a partner?"[1]. The task is a cooperation work that take place in front of a whiteboard by two persons. In our method, we focus on a top-view image for the estimation. Figure 1 shows an example of an image from an overhead camera. We extract the standing location, operation ratio and head direction of each person from the top-view image. We analyze the relation between the cooperation level and each feature. Then, we apply the features to two machine learning approaches: AdaBooost and multiple linear regression. The goal of our study is to provide the cooperation level estimated by the features to teachers in the PBL.

In the next section, we explain related work. Next, we describe our method in Section 3. In the section, we discuss nonverbal information for the cooperation estimation first, and then describe features and classifiers based on machine learning approaches for the task. In Section 4, we discuss our experimental results in terms of the features and classifiers. Finally, we conclude our methods in Section 5.

2 Related work

Many researchers have studied methods for estimating atmosphere and participant's mind in an interaction. Takashima et al. [15] have reported an analysis

[1] Note that we are not concerned with the quality of the output of each pair work in this paper.

of nonverbal cues and atmosphere in six person conversations. The nonverbal information was acquired with some sensor devices such as a 3D tracker, an acceleration sensor and a tablet device. Mota and Picard [9] have proposed a posture recognition method for a person seated in a chair. They used a leap chair with pressure sensors. In general, using particular devices is, however, costly and cumbersome for participants.

One solution for the issue is to utilize cameras for the extraction of nonverbal information. Nakamura et al. [10] have proposed a method for estimating learners' subjective impressions for an e-learning system. They used facial information, facial expressions, gaze and head poses from a stereo-camera. Jayagopi et al. [4] have proposed a method for mining and validating group speaking and gaze patterns. They captured images and speech from two web-cameras and a commercial array microphone. Grafsgaard et al. [2] have proposed a method for analyzing posture and affect for intelligent tutoring systems. They captured depth information with a Kinect. These studied, however, treated a e-learning system for one person and persons seated in chairs. Our research target is a PBL environment in which participants have actions.

In this paper, we use a depth camera for the extraction of nonverbal information. The task is pair work. A participant might be occluded by another participant if a camera is placed in front of participants. To solve this problem, we apply an overhead camera to our task. By using the overhead camera, the problem of occluded images is solved. We have reported the effectiveness of the use of the overhead camera for a person identification task [6, 11] and a posture identification task [5]. In addition, the method with the overhead camera has the advantage that psychological resistance is reduced because the camera does not capture the face image. Furthermore, the restriction of the location of a camera is reduced because the camera does not need to capture the person's face.

3 Proposed method

In our method, we use Microsoft Kinect[2] to capture pair work activities. The Kinect camera can handle depth information. We extract several features from the captured images. Finally, we estimate the cooperation level in each pair work on the basis of these features.

3.1 Nonverbal information

In this paper, we focus on nonverbal information for estimating the cooperation level. The nonverbal information in this paper is divided into three categories: (1) standing location, (2) operation ratio and (3) head direction.

First, we record a pair work activity with Kinect. Our method detects head areas of each person by using depth information from the Kinect. Next it computes the centroid of each head area. We regard the centroid as the standing location of a person. Figure 2 shows an example of the process.

[2] http://www.xbox.com/en-US/Kinect

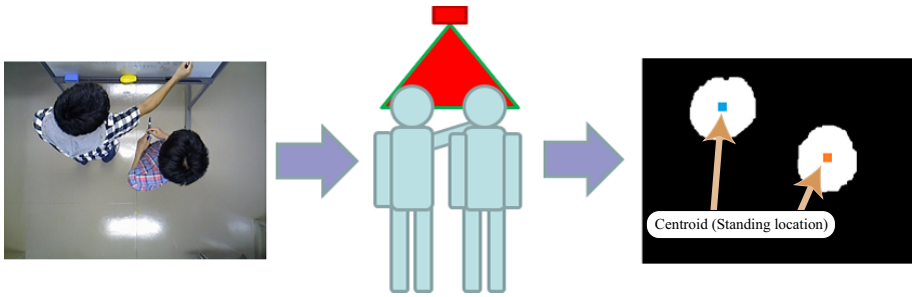

Fig. 2. The person area extraction and standing location.

The second nonverbal information is an operation ratio of each person. Here the operation denotes a pointing gesture and a writing activity on the whiteboard. As a preliminary experiment, we generated a classifier for identifying the operations in a video. We used the locations of a pen tip and a fingertip by using depth information as the features for the classifier. In addition, we used the dimensions of person's area that are near the whiteboard and the distance between the centroids of the person area and head area as the features. Figure 3 shows an example of the process. We evaluated a machine learning technique using these feature with 300 images. The accuracy was 80%. We think that it is not sufficient to investigate the cooperation level in pair work although the accuracy is not low. Therefore, we prepare annotated data for the operation ratio. We manually annotate each image with three classes; pointing, writing and otherwise.

The third nonverbal information is a head direction. We divide an image to six directions. Figure 4 shows the six directions. We assign five directions to the front of the head and one direction to the backward. We also estimated the direction in each image automatically on the basis of the head shape. However, the accuracy was 47%[3]. Therefore, we also prepare manual annotated data for the head direction.

3.2 Features

We explained the nonverbal information that we use in the previous section. In this section, we describe features for the cooperation level estimation in detail. We introduce eight features to the estimation approach. These features are extracted from the output of the previous section.

Location deviation We compute the standard deviation of the standing location. A large location deviation value denotes that the person is active during the pair work.

[3] The reasons why the accuracy was extremely low that were (1) a person sometimes stood outside of the camera range and (2) a head shape depended on the standing location.

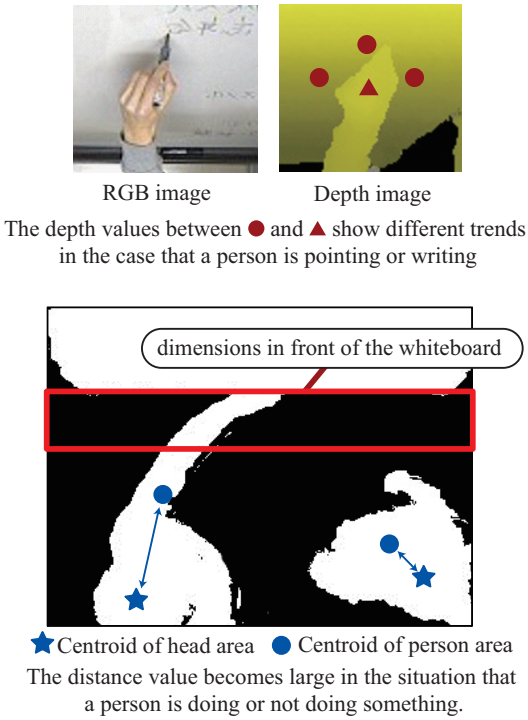

Fig. 3. The features for the operation identification.

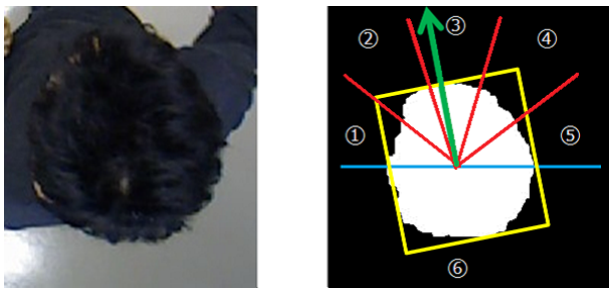

Fig. 4. Head direction.

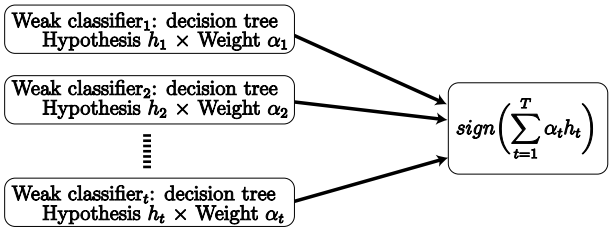

Fig. 5. The AdaBoost algorithm.

Center ratio We compute a center ratio of two person. This ratio is which person stands to the center area in long time.

Average distance We compute an average distance value between two persons from all image frames.

Operation ratio We compute an operation ratio. Here the target operations are (1) pointing and (2) writing. This feature is the ratio of these two operations in all frames.

Head turn This is a difference of the head direction between two frames. If the head direction in the current frame differs from that in the previous frame, our method counts up the number of the changes.

Gaze ratio We measure the frequency of the same head direction that occurs in n consecutive frames. For example, we set n to 3. If we obtain a direction sequence, ③ → ③ → ③ → ③ → ③ → ③ → ② → ② → ②, the frequency is 2. In this paper, we use three types of n; $n = 3, 5, 10$.

Eye direction We assume that the head direction implies the gaze direction of eyes. We compute the ratios of six directions in Figure 4 in all frames.

Partner gaze ratio We compute the gaze ratio to a partner. We compute the ratios of the partner side (④ and ⑤) and the opposite side (①, ② and ③) for a person in the left side. We also compute the ratios of the partner side (① and ②) and the opposite side (③, ④ and ⑤) for a person in the left side

3.3 Classifier

On the basis of the extracted features, we generate a classifier for the cooperation level estimation. In this paper, we employ the AdaBoost algorithm [1] as the classifier. The AdaBoost algorithm is one of the most famous machine learning techniques. It generates a strong classifier by combining some weak classifiers. In this paper, we implement the AdaBoost with the open source software Weka[17]. We use the C4.5 algorithm [13] as the weak classifiers[4]. C4.5 is also one of the famous machine learning techniques, which generates a decision tree. Figure 5 shows the outline of the AdaBoost algorithm.

In addition, we introduce another approach. We employ the multiple linear regression analysis with the stepwise method for the estimation.

[4] Actually, it is "48" in Weka.

4 Experiment

We evaluate our method with an annotated data set of pair work. In this section, we explain the experimental settings first. Then, we consider the relevancy between the cooperation level and some features. Finally, we discuss the results of two machine learning approaches.

4.1 Setting

We collected pair work with a whiteboard by using an overhead camera. The pair work consisted of two processes. First, each test subject wrote words related to a given topic to the whiteboard. In this experiment the topics were "Summer" and "Autumn". For example, the written words were "Fireworks", "Festival", "Summer holiday" and so on. Next, they classified the written words on the whiteboard into several categories in a subjective manner. For example, a pair categorized these three words, "Fireworks", "Festival", "Summer holiday", as "Summer event". The categorization process depended on the free discussion of each pair. Each operation time of two process, namely the listing of words and categorization, was three minutes, respectively.

The number of test subjects was 16 persons. We generated ten groups from them. Five groups have the acquaintance relationship and others were the first meet. The cooperation level of each pair was determined by one annotator[5]. The range of cooperation level was 1 (bad) to 5 (good).

We obtained 1800 frames from Kinect. We extracted 180 images from them with respect to each ten frame. We divided 180 images into 90 images as the anterior half and 90 images as the posterior half in each pair work.

4.2 Result

In this section, we discuss three main nonverbal aspects described in Section 3.1, namely standing location, operation ratio and head direction, in the experimental data first. Then, we discuss the accuracy of the AdaBoost and the reliability of the multiple linear regression approach.

Discussion on location, operation and head direction
Table 1 and Table 2 show the experimental data about acquaintance groups and first-meet groups, respectively. In the tables, CL denotes the cooperation level by one annotator. LocDev, CentRatio, AveDist and OpRatio are the location deviation, the center ratio, the average distance and the operation ratio in Section 3.2, respectively. "A" to "J" denote the group ID. "(M)" and "(F)" (F)"

[5] Actually, we collected a questionnaire about the cooperation level of the pair work from the test subjects. However, there was a large difference between self-evaluation and cooperation level by the annotator. In other words, the self-evaluation differ from the actual cooperation level. We think that the reason was that each test subject was liable to pay mind to the partner in the pair work.

Table 1. Result of Group A to E.

Acquaintance	CL	LocDev	CentRatio	AveDist	OpRatio
A1(M)	5/5	31.97/37.80	1.4/1.2	356.2/354.7	44.4/41.1
A2(M)	5/5	32.13/53.12	98.6/98.2		62.2/81.1
B1(M)	5/5	35.68/34.98	30.0/26.7	395.5/387.3	47.8/60.0
B2(F)	5/5	26.07/39.17	70.0/73.3		57.8/40.0
C1(M)	3/1	13.35/17.07	5.6/0.0	429.3/446.5	27.8/5.6
C2(F)	4/2	41.56/23.95	94.4/100.0		38.9/42.2
D1(M)	4/5	42.45/46.63	0.0/16.7	356.6/343.2	12.2/58.9
D2(M)	4/5	26.29/45.35	100/83.3		54.4/47.8
E1(F)	4/3	18.42/15.08	15.6/14.4	266.5/270.8	35.5/15.5
E2(F)	5/4	33.00/24.37	84.4/85.6		62.2/50.0

Table 2. Result of Group F to J.

First-meet	CL	LocDev	CentRatio	AveDist	OpRatio
F1(M)	4/3	22.97/33.68	59.3/100.0	395.8/358.6	64.4/23.7
F2(F)	3/2	18.22/19.12	40.7/0.0		6.8/18.6
G1(M)	4/2	20.72/20.71	2.1/68.8	328.7/339.7	41.7/6.25
G2(F)	4/4	14.91/36.02	97.9/31.2		47.9/54.1
H1(M)	3/3	9.96/21.58	0.0/1,9	312.5/308.4	28.8/22.6
H2(M)	4/3	21.22/18.12	100.0/98.1		35.8/28.8
I1(M)	4/4	22.65/31.25	0.0/0.0	353.0/341.3	55.6/44.4
I2(F)	4/4	20.44/15.94	100/100.0		51.1/42.2
J1(M)	4/1	30.48/16.42	82.2/100.0	370.8/356.3	13.3/2.2
J2(F)	4/1	23.39/14.11	17.8/0.0		44.4/20.0

are male and female. "/" is the border of the anterior half and the posterior half about each feature. For example, the test subject A1 was male and the anterior half and the posterior half features of his location deviation were 31.97 and 37.80.

First, we discuss the location features. The LocDev became larger in the situation that the cooperation level was high, e.g., Group A and B. On the other hand, the cooperation levels about the groups or persons with the small LocDev value, such as Group J and person C1, were low. In other words, persons with high cooperation level were active during the pair work in this experiment. This denotes that large and frequent motions of the head and body are generated in the case that the pair work proceeds smoothly. Furthermore, the CentRatio of persons with leadership potential tended to increase the value, such as A1 and C2. The reason was that such persons were enterprising and willing, e.g.,

Table 3. Confusion matrix.

	Estimated		
	Cooperative	Neither	Uncooperative
Cooperative	25	1	1
Neither	1	2	3
Uncooperative	1	2	4

writing on whiteboard. In addition, for a female pair, Group E, the AveDist had a tendency to become small.

Next, we discuss the operation ratio. The OpRatio became larger for the high cooperation level, such as Group A and B, and lower for the low cooperation level, such as Group J. This is the same tendency as the LocDev. In contrast, there was no significant difference between acquaintance and first-meet about each features.

Finally, we discuss the head direction. For the head direction features, there was no relativity about the cooperation level. The groups of the high cooperation level had a tendency to increase operation time about writing on whiteboard. Therefore, the persons of the group fixed their eyes to the same direction. On the other hand, the groups of the low cooperation level had a tendency to increase non-operation time because they stood by doing nothing. Therefore, the persons of the group also fixed their eyes to the same direction, namely the white board. As a result, there was no difference between the high and low cooperation groups.

Discussion on classifiers

Next, we evaluated two machine learning approaches with our features. The annotated data consisted of 20 persons for 10 groups. We divided the data of 20 persons into two parts; the anterior half and the posterior half. Hence we obtained 40 instances as the experimental data. We generated a three-class problem from the five cooperation level. In other words, we integrated the cooperation level 4 and 5 to "Cooperative" and 1 and 2 to"Uncooperative". The cooperation level 3 was class "Neither"

We evaluated the AdaBoost with 20-fold cross validation. On the basis of a preliminary experiment, we used the location deviation and the operation ratio as the features of the AdaBoost. The accuracy rate was 77.5%. Table 3 shows the confusion matrix of the experimental result. Although the accuracy was relatively-good, the evaluation data was unbalance. Most instances were the class "Cooperative". To validate the effectiveness of our method, we need to acquire more pair work data. The reason why the class "Cooperative" became the great majority was that we applied pair work to the task. Pair work has a natural tendency to cooperate with each other because the group consists of only two persons. Extension to a multi-party task is the most important future work.

Finally, we evaluate the reliability of the multiple linear regression approach. For the task, we used the original five cooperation level in the data. We applied

all nonverbal features to the regression. In the analysis, we select the optimal features from them by using the stepwise method. As a result, LocDev and OpRatio were selected as explanatory variables. The equation about the cooperation level (CL) is as follows:

$$CL = 1.256 + 0.0415 \times LocDev + 0.0353 \times OpRatio \quad (1)$$

The standardized partial regression coefficients of LocDev and OpRatio were 0.371 ($p < .005$) and 0.572 ($p < .001$), respectively. The adjusted R^2 was 0.649. The nonverbal features were effective because the $AdjR^2$ was very high. However, the data was annotated by one annotator. Therefore, we need to verify the annotated cooperation level by several annotators. This is also the important future work.

5 Conclusions

In this paper, we discussed the estimation of a cooperation level in pair work. The task was the cooperation work that take place in front of a whiteboard by two persons. This study leads to one useful task for education support systems and project based learning.

We explained several nonverbal features and analyzed real data by them. Some knowledge was acquired about relations between the cooperation level and the nonverbal information, such as the location deviation and operation ratio. On the other hand, the head direction features were not effective for the estimation of the cooperation level. We also applied the features to two machine learning approaches. The AdaBoost with the features produced 77.5% as the accuracy. The adjusted R^2 of the multiple linear regression was 0.649. However, it is important to investigate and incorporate new nonverbal features to our method for the improvement of the cooperation level estimation.

Future work includes (1) evaluation of the method with a large-scale dataset, (2) utilization of other nonverbal features, (3) improvement of the operation identification and the head direction estimation, (4) annotation of the cooperation level by several annotators and (5) extension to a multi-party task (cooperation by 3 or more persons).

References

1. Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. In *Proceedings of International Conference on Machine Learning*, pages 148–156, 1996.
2. Joseph F. Grafsgaard, Kristy Elizabeth Boyer, Eric N. Wiebe, and James C. Lester. Analyzing posture and affect in task-oriented tutoring. In *Proceedings of FLAIRS Conference 2012*, 2012.
3. Cindy E. Hmelo-Silver. Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16:235–266, 2004.

4. Dinesh Babu Jayagopi, Dairazalia Sanchez-Cortes, Kazuhiro Otsuka, Junji Yamato, and Daniel Gatica-Perez. Linking speaking and looking behavior patterns with group composition, perception, and performance. In *Proceedings of the International Conference on Multimodal Interaction (ICMI), Santa Monica, USA*, 2012.
5. Kazuaki Komatsu, Kazutaka Shimada, and Tsutomu Endo. Posture identification for evaluation of multi-party interaction. In *Technical report of IEICE, HCS*, volume 112, pages 25–30, 2012.
6. Daichi Kouno, Kazutaka Shimada, and Tsutomu Endo. Person identification using top-view image with depth information. In *Proceedings of 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD 2012)*, pages 140–145, 2012.
7. Shiro Kumano, Kazuhiro Otsuka, Masafumi Matsuda, and Junji Yamato. Understanding empathy/antipathy perceived by external observers based on behavioral coordination and response time. In *Proceedings of HCG symposium*, 2012.
8. Marwa Mahmoud, Tadas Baltrusaitis, Peter Robinson, and Laurel Riek. 3d corpus of spontaneous complex mental states. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII 2011)*, Lecture Notes in Computer Science, 2011.
9. Selene Mota and Rosalind W. Picard. Automated posture analysis for detecting learner ' s interest level. In *Proceedings of Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI*, 2003.
10. Kazuaki Nakamura, Koh Kakusho, Masayuki Murakami, and Michihiko Minoh. Estimating learners' subjective impressions of the difficulty of course materials in e-learning environments. In *Proceedings of APRU 9th Distance Learning and Internet Canference 2008*, pages 199–206, 2008.
11. Ryota Nakatani, Daichi Kouno, Kazutaka Shimada, and Tsutomu Endo. A person identification method using a top-view head image from an overhead camera. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 16(5):696–703, 2012.
12. Mai Omura and Kazutaka Shimada. Estimation of subjective impressions of difficulty in quiz dialogue. In *Technical report of IEICE, NLC2013-2*, pages 7–14, 2013.
13. J. R. Quinlan. *C4.5 Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
14. Kazutaka Shimada, Akihiro Kusumoto, Takahiko Yokoyama, and Tsutomu Endo. Hot spot detection in multi-party conversation using laughing feature. In *Technical report of IEICE, NLC2012-7*, pages 25–30, 2012.
15. Kazuki Takashima, Kazuyuki Fujita, Hitomi Yokoyama, Yuichi Itoh, and Yoshifumi Kitamura. A study of nonverbal cues and subjective atmosphere in six person conversations. In *Proceedings of IEICE SIG-HCS*, volume 112, pages 49–54, 2012.
16. Marjorie Fink Vargas. *Louder than Words: an Introduction to Nonverbal Communication*. Iowa State Press, 1986.
17. Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, third edition, 2011.
18. Takuya Yamane, Kazuaki Nakamura, Mayumi Ueda, Masayuki Mukunoki, and Michihiko Minoh. Detection of interaction between a lecturer and learners based on their actions. In *Proceedings of The Japanese Society for Artificial Intelligence, SIG-ALST(Advanced Learning Science and Technology)*, 2010.

19. Takahiko Yokoyama, Kazutaka Shimada, and Tsutomu Endo. Hot spot detection in multi-pary conversation using linguistic and non-linguistic information. In *Proceedings of NLP2012 (in Japanese)*, 2012.