Interspeech 2022
18-22 September 2022, Incheon, Korea

Analysis of praising skills focusing on utterance contents

Asahi Ogushi[1], *Toshiki Onishi*[1], *Yohei Tahara*[2], *Ryo Ishii*[3],
Atsushi Fukayama[3], *Takao Nakamura*[3], *Akihiro Miyata*[2]

[1]Graduate School of Integrated Basic Sciences, Nihon University
[2]College of Humanities and Sciences, Nihon University
[3]NTT Human Informatics Laboratories, NTT Corporation

chas22009@g.nihon-u.ac.jp, chto22002@g.nihon-u.ac.jp, chyo19046@g.nihon-u.ac.jp,
ryo.ishii.ct@hco.ntt.co.jp, atsushi.fukayama.ap@hco.ntt.co.jp,
takao.nakamura.vp@hco.ntt.co.jp, miyata.akihiro@acm.org

Abstract

Praising behavior is considered an important method of communication. It is considered effective in building good relationships with others and bringing out the best in them. However, there are quite a few people who have difficulty in praising successfully. These people have difficulty judging and improving their praising skills. The reason is that they have not clarified which behaviors are important for successfully praising. In this paper, we analyze praising behavior by focusing on the content of utterances in Japanese. We construct machine learning models for estimating praising skills to analyze behaviors for successfully praising. Then, we extract features from the utterances of persons who give praise (praiser) and those who receive praise (receiver). The results showed that the best performance of our models was $F1 = 0.632$. We analyzed praising skills from the results of that model. An utterance of praising scene, one receiver utterance immediately preceding the utterance of praising scene, and four receiver utterances immediately following it are crucial for estimating praising skills. The cosine similarity between an utterance by a praiser and one by a receiver is also important as a feature. Thus, we confirmed that the utterances of receiver are similar to the utterances of praiser.

Index Terms: multimodal interaction, verbal behavior, praise

1. Introduction

Praising behavior is considered an important method of communication in daily life and social activities. However, some people would like to become better at praising, because they are not confident in their ability to do it. To improve their praising skills, they need to understand their own praising skills. However, they have difficulty understanding the degree of their skills. To learn about their own skills, they need to ask others, such as family and friends, to evaluate them and to receive guidance from experts. Therefore, we attempt to clarify which behaviors are important for successfully praising a partner during a dialogue. Onishi et al. [1][2] have worked to clarify the behaviors for successfully praising from nonverbal behaviors such as head motion, facial expressions, and voice behavior. However, verbal behavior is also considered important for praising. In this study, we develop machine learning models that use the linguistic features of the speaker to estimate praising skills, thereby clarifying which features are important for the praiser and receiver. The main contributions of this paper are as follows.

- To clarify whether linguistic features can be used to estimate praising skills.
- To clarify which utterances of the praiser and receiver are important for estimating praising skills.

2. Related work

2.1. Studies on praising behavior

Praising behavior has been studied in the fields of psychology and sociology. In educational psychology, particular emphasis is placed on the teacher-student relationship and getting the most out of students' abilities[3][4]. Praising is verbal and nonverbal behavior that expresses admiration directed toward the actions and character of the target and is considered a complex form of social communication rather than one-way communication from a person giving praise to a receiver [5][6][7]. Morton et al. [8] modeled the sequence of praising actions and defined what factors and responses are at work in the act of praise. They believe that the objects of praise are the appearance, possessions, accomplishments, and efforts of the person being praised. It was suggested that the praiser forms their compliments by listening to the "accomplishments" and "effort" of the person to be praised. The praiser's response is thought to be contained in the words that the person given praise says after the praise.

2.2. Estimation of personality traits and performance

Many studies have analyzed behaviors and abilities using human verbal and nonverbal behaviors, such as communication skills, presentation skills, empathy, self-disclosure, and speaker persuasion. This study is closely related to studies that use human verbal and nonverbal information to analyze behavior and abilities in specific tasks and situations. Okada et al. [9] constructed a regression model to predict communication skill scores using turn-taking , prosodic information, head movements, parts of speech, and dialogue acts as features based on ratings of communication skills by experienced personnel managers. Tan et al. [10] used verbal and nonverbal behaviors in dialogue to construct and evaluate a model for predicting whether listeners empathize with the speech of others, and they reported that speech and speech content, especially the speaker's speech content, are important in predicting the degree of empathy of the listener. Perez-Rosas et al. [11] analyzed speaker emotions in online videos using verbal and nonverbal information. They report that the highest performance in emotion estimation was obtained when using audio, visual, and speech content modalities. Soleymani et al. [12] showed that the content of verbal behavior is related to self-disclosure through a correlation analysis of verbal and nonverbal behavior during dialogue. They also constructed a multimodal deep neural network to estimate

10.21437/Interspeech.2022-11200

the level of self-disclosure. The results show that a unimodal model using only verbal behavior can most accurately estimate the degree of self-disclosure. Onishi et al. [1][2] reported that nonverbal behaviors related to head, face, and voice are useful for estimating the ability to praise. Therefore, we attempt to estimate praise using not only nonverbal behaviors but also speaker utterances.

3. Research goals

As mentioned in 2.1, many studies on praising behavior have been reported in psychology and sociology. There has been little engineering analysis of how to behave in actual dialogue situations. In addition, as mentioned in 2.2, many studies have been reported that analyzed behaviors and abilities in specific tasks and scenes in terms of verbal and nonverbal behaviors. These studies analyzed speakers' personality traits and abilities such as presentation, empathy, and self-disclosure in communication. However, few studies have analyzed praising behavior using human verbal and nonverbal behaviors. Therefore, we analyzes praising behavior from an engineering perspective and attempts to clarify which behaviors are important in praising behavior. Onishi et al. [1][2] analyzed praising skills using head motion, facial expression, and voice behavior. However, the speaker's utterances appear crucial in praising behavior[8]. Many studies have reported similar levels of estimation performance using only nonverbal and verbal behaviors when evaluating a specific task or human behavior. From 2.1, the praising behavior is a complex communication between the praiser and receiver. This indicates that the utterance of praising scene is closely related to the receiver's utterance. Therefore, we focus on the utterance of both the praiser and receiver in praising behavior. To praise successfully, we need to clarify the relationship between the utterance of a praiser and a receiver. Based on the above, our research goals are the following three items.

Goal 1. Predicting praising skills from a speaker's utterances.

Goal 2. Analyze linguistic features that can predict praising skills with high performance.

Goal 3. Analyze which utterances of a speaker should be used to predict praising skills with high performance.

4. Dialogue corpus

In this paper, we used a corpus of dialogues developed by Onishi et al. [1][2]. This corpus contains face-to-face two-party dialogues annotated with dialogue data and includes an evaluation of praising skills. The details of the corpus are described in this section.

4.1. Recording of two-party dialogue

Onishi et al. [1][2] recorded two-party dialogues in Japanese to record verbal and nonverbal behaviors. The participants in the two-party dialogues were 34 university students in their twenties (28 males and 6 females) who were divided into 17 pairs. Among the 17 pairs, 14 pairs included participants meeting for the first time, 2 pairs included acquaintances, and 1 pair included friends. To begin recording dialogues, they requested participants prepare two or more episodes about things that they had been working hard on to prepare materials for the dialogues. The participants were seated facing each other and separated by 180 cm, as shown in Figure 1. The dialogues were recorded using a video camera to record each participant's head and face behaviors and a microphone to record each participant's voice.

Figure 1: *Two-party face-to-face dialogue.*

	Scenes	Mean (sec)	Max (sec)	Min (sec)
Utterance (praiser)	2701	1.324	23.117	0.062
Utterance (receiver)	3413	2.040	26.234	0.018
Praising scenes	228	2.018	9.127	0.368

Table 1: *Information on utterance and praising scenes.*

Each pair of participants (participants A and B) performed dialogues (1) to (3) in accordance with the experimenter's instructions.

(1) A self-introduction (5 min).

(2) Dialogue with participant A as a praiser and participant B as a receiver (5 min).

(3) Dialogue with participant B as a praiser and participant A as a receiver (5 min).

They recorded 17 pairs of dialogues (1) to (3) for a total of 255 minutes of two-party dialogues. Dialogue (1) (self-introduction) was not used in their analysis because many of the pairs were meeting for the first time, and its purpose was simply to relieve the tension between participants. In dialogues (2) and (3), the receiver was instructed to discuss the things that they had been working hard to accomplish. To ensure that the participants conversed naturally regarding a variety of topics, they also allowed them to discuss topics that they had not prepared beforehand. The praiser was instructed to praise the receiver. However, they allowed the participants to raise questions and react freely to avoid any unnatural dialogues that would have involved unilateral praising. This procedure was approved by the ethics committee.

4.2. Annotation of dialogue data and evaluation of praising skills

Onishi et al. [1][2] annotated the dialogue data recorded in 4.1. They used ELAN [13], a tool for annotating video and audio data, to manually annotate utterance scenes in the video and audio data of each participant. The results of the scenes are presented in Table 1.

Utterance scene: continuous voice intervals with a silent interval of less than 400 ms.

The interval between consecutive utterances was set to less than 400 ms so that the intervals of utterances were natural.

The evaluation of praising skills was conducted by five third-party annotators who did not participate in the two-party dialogues. Specifically, annotators made the following judgment and evaluation for each utterance scene featuring the praiser extracted in 4.2, referring to the video data recorded from the video camera set up in front of the praiser and the audio data recorded from the microphone attached to the praiser.

- Judgment of whether the praiser praises the dialogue partner in the scene.
- If the praiser did praise the dialogue partner, an evaluation of praising skills on a 7-point Likert scale from 1 (I do not think the praiser is successfully praising) to 7 (I think the praiser is successfully praising).

From the above, we defined praising scenes and praising scores. Information on the praising scenes is shown in Table 1.

Praising scene: a scene in which three or more annotators judged the praiser to be praising in each utterance scene.

Praising score: the mean of the evaluations by annotators who judged the praiser to praising in each praising scene.

They evaluated the rate of concordance of praising scores among annotators using intraclass correlation coefficients (ICC) [14]. They calculated the ICC for each combination of three to five annotators, and calculated the weighted average by considering the number of samples. The results were $ICC(2,k) = 0.571$. This suggests that the praising scores were reliable data with a medium level of concordance among annotators.

5. Analysis of features that contribute to praising skills

5.1. Feature extraction

We extracted the linguistic features of the speaker's utterances from the range shown in Figure 2. We used the utterances of praising scenes (hereafter called "Ps") as the utterances by the praisers. We used ten utterances by the receiver, including the five utterances immediately before a Ps (hereafter called "B1 to B5") and five utterances immediately after the Ps (hereafter called "A1 to A5"). To extract linguistic features from the speaker's utterances, we used BERT [15], which represents a speaker's utterances as vectors. BERT is a language model that can train on large unsupervised data sets. It is a multi-layer bidirectional transformation network and is effective for representing word embeddings. In this paper, we transformed each utterance into a 768–dimensional feature vector using a BERT model pre-trained using Japanese Wikipedia. As a specific extraction method, we obtained an embedded representation of [CLS], a special token at the top of the final layer of the BERT process for an input sentence.

In addition, the utterance of a praiser and that of a receiver are closely related [8]. Therefore, we considered the similarity between them to be important. We calculated the cosine similarity of each utterance vector of the praiser and receiver and used it as a feature. This allowed us to clarify whether the similarity between the utterances is necessary.

5.2. Estimation of praising scores

We developed machine learning models that use the linguistic features of the speaker to estimate praising skills. On the basis of the above, we divided the praising scenes (228 scenes in total) into three groups: low, medium, and high group. We developed a classifier that estimates which class a praising score belongs to on the basis of linguistic features. To keep the number of praising scenes in each group as equal as possible, the praising scores for the low to high groups were defined as follows. Threshold values were defined by making the number of scenes in the three groups be as equal as possible and by making sure that scenes with the same praising scores did not exist multiple groups [1].

Low group: scenes in which utterance of praising scene with a praising score of 3.8 points or less (82 scenes in total).

Middle group: utterance of praising scene with a praising score greater than 3.8 points and less than 4.4 points (65 scenes in total).

High group: utterance of praising scene with a praising score of 4.4 points or higher (81 scenes in total).

We used random forests [16], which can evaluate the importance of features, to develop an estimation model. We tuned hyperparameters such as the learning rate and tree depth using Hyperopt [17]. Feature selection was repeated until the model stopped improving by removing the least important features sequentially. The dataset was randomly divided into 90% training data and 10% test data. The task of estimating the class to which the test data belongs using a model trained on the training data was repeated 100 times. Machine learning models were built by varying the scope and method of feature extraction.

5.3. Results of the proposed models

Table 2 shows the estimation results for the constructed machine learning models. The baseline was the chance level. The model with the highest performance used an utterance of praising scene, one receiver utterance immediately before the utterance of praising scene, and four receiver utterances immediately after the utterance of praising scene. The result of this model was $F1 = 0.632$.

5.4. Discussion

Based on the results in Section 5.3, we discuss the relationship between speakers' utterances and praising skills.

Goal 1. Predicting praising skills from a speaker's utterances. Based on the results in section 5.3, the best performing model (M25, $F1 = 0.632$) had a higher accuracy than the chance level (Baseline, $F1 = 0.357$). Furthermore, the performance of M25 exceeded that of the machine learning model for estimating praising skills based on nonverbal behaviors ($F1 = 0.548$) [2]. This allows us to confirm that estimating praising skills using the content of the speaker's utterances is effective. However, of course, the act of praise comprises verbal and nonverbal behaviors. In the future, we should analyze praising skills using both verbal and nonverbal behaviors.

Goal 2. Analyze linguistic features that can predict praising skills with high performance. The model with the highest performance (M25) was based on the BERT vector of the utterance of praising scene and the cosine similarity between the praiser's utterance and the receiver's utterance. Many of the statements from the high group did not merely include words of praise, but also specific reasons for praising. This allowed a variety of concepts to be included in successful praising utterances. The BERT vector probably represents the diversity of concepts. Related to the above, there were many cases in the high group where the praiser and the receiver specifically referred to each other's statements. This made the concept of the

[1] The number of scenes in each group should be equal, but it was not equal because many scenes had the same score.

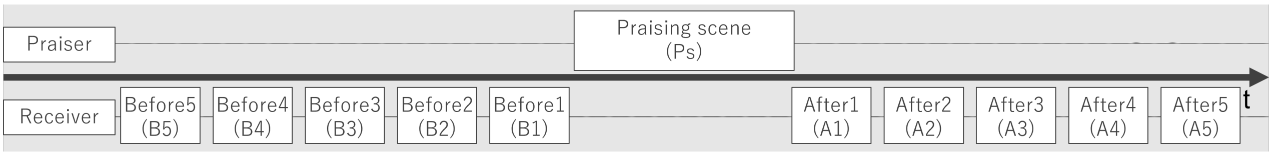

Figure 2: *Speaker's utterance with linguistic features extracted.*

	Ps	B5	B4	B3	B2	B1	A1	A2	A3	A4	A5	F1 (praiser + receiver)	F1 (praiser + cosine similarity)	F1 (praiser + receiver + cosine similarity)
Baseline												0.357	0.357	0.357
M0	✓											0.620	0.620	0.620
M1	✓					✓						0.593	0.584	0.580
M2	✓				✓	✓						0.589	0.601	0.571
M3	✓			✓	✓	✓						0.565	0.595	0.575
M4	✓		✓	✓	✓	✓						0.578	0.613	0.568
M5	✓	✓	✓	✓	✓	✓						0.568	0.597	0.560
M6	✓						✓					0.616	0.593	0.619
M7	✓					✓	✓					0.598	0.602	0.589
M8	✓				✓	✓	✓					0.561	0.618	0.584
M9	✓			✓	✓	✓	✓					0.574	0.612	0.568
M10	✓		✓	✓	✓	✓	✓					0.592	0.576	0.580
M11	✓	✓	✓	✓	✓	✓	✓					0.599	0.609	0.562
M12	✓						✓	✓				0.607	0.611	0.582
M13	✓					✓	✓	✓				0.587	0.595	0.578
M14	✓				✓	✓	✓	✓				0.587	0.614	0.572
M15	✓			✓	✓	✓	✓	✓				0.571	0.611	0.562
M16	✓		✓	✓	✓	✓	✓	✓				0.569	0.574	0.575
M17	✓	✓	✓	✓	✓	✓	✓	✓				0.583	0.608	0.559
M18	✓						✓	✓	✓			0.563	0.586	0.579
M19	✓					✓	✓	✓	✓			0.563	0.604	0.550
M20	✓				✓	✓	✓	✓	✓			0.580	0.580	0.576
M21	✓			✓	✓	✓	✓	✓	✓			0.580	0.604	0.575
M22	✓		✓	✓	✓	✓	✓	✓	✓			0.584	0.560	0.592
M23	✓	✓	✓	✓	✓	✓	✓	✓	✓			0.539	0.608	0.577
M24	✓						✓	✓	✓	✓		0.613	0.595	0.569
M25	✓					✓	✓	✓	✓	✓		0.594	**0.632**	0.595
M26	✓				✓	✓	✓	✓	✓	✓		0.584	0.610	0.577
M27	✓			✓	✓	✓	✓	✓	✓	✓		0.549	0.593	0.562
M28	✓		✓	✓	✓	✓	✓	✓	✓	✓		0.566	0.572	0.577
M29	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		0.582	0.574	0.556
M30	✓						✓	✓	✓	✓	✓	0.566	0.597	0.549
M31	✓					✓	✓	✓	✓	✓	✓	0.560	0.565	0.550
M32	✓				✓	✓	✓	✓	✓	✓	✓	0.550	0.569	0.567
M33	✓			✓	✓	✓	✓	✓	✓	✓	✓	0.580	0.592	0.572
M34	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	0.550	0.624	0.600
M35	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.581	0.613	0.555

Table 2: *Evaluation results for each model with varying feature extraction ranges and feature extraction methods (praiser: vector representation of Ps' statements, receiver: vector representation of B1 to B5 and A1 to A5 statements, cosine similarity: cosine similarity of Ps' statements and B1 to B5 and A1 to A5 statements).*

praiser's utterance and the receiver's utterance similar. The cosine similarity probably represents the proximity of the concept.

Goal 3. Analyze which utterances of a speaker should be used to predict praising skills with high performance. The range of features extracted for the highest performing model (M25) had praising scene (Ps), an utterance of the receiver immediately before the praising scene (B1), and four utterances of the receiver immediately after the praising scene (A1 to A4). The receiver's utterances immediately before the praiser's utterance seem to contain the information to be praised. The receiver's utterances immediately after the praiser's utterance are considered to indicate the level of satisfaction with praise. For the above reasons, we consider that M25 involving both before and after utterances yielded the best performance.

This study has some limitations. First, we have not considered the utterances of receivers that overlap with praising scenes. These utterances may be more important than other utterances. In the future, we would like to analyze these utterances taking into account. Second, the speakers were exclusively Japanese. If the language, historical background, and culture are different, the results might be different. In the future, we would like to analyze praising skills taking into account differences in language or culture.

6. Conclusion

In this paper, we analyzed how the utterance contents of the praiser and receiver in Japanese is related to praising skills. The results showed that these skills can be predicted from the content of utterances. In addition, we clarified ranges of utterances from which to extract appropriate linguistic features to predict praising skills. The ranges of significant utterances are as follows: the utterance in the praising scene, one utterance by the receiver immediately before the praising scene, and four utterances by the receiver immediately after the praising scene. Thus, predicting praising skills requires attention to the content of both the praiser's and receiver's utterance. In the future, we intend to conduct a multimodal analysis by adding nonverbal behaviors such as facial expressions and voice behavior.

7. References

[1] T. Onishi, A. Yamauchi, R. Ishii, Y. Aono, and A. Miyata, "Analyzing nonverbal behaviors along with praising," in *Proceedings of the 22nd ACM International Conference on Multimodal Interaction (ICMI'20)*, 2020, pp. 609–613.

[2] T. Onishi, A. Yamauchi, A. Ogushi, R. Ishii, A. Fukayama, T. Nakamura, and A. Miyata, "Modeling japanese praising behavior by analyzing audio and visual behaviors," *Frontiers in Computer Science*, vol. 4, 2022.

[3] V. M. Catano, "Relation of improved performance through verbal praise to source of praise," *Perceptual and Motor Skills*, vol. 41, no. 1, pp. 71–74, 1975.

[4] J. Brophy, "Teacher praise: Afullctional analysis," *Review of Educational Research*, vol. 51, no. 1, pp. 5–32, 1981.

[5] J. Henderlong and M. R. Lepper, "The effects of praise on children's intrinsic motivation: A review and synthesis," *Psychological Bulletin*, vol. 128, no. 5, pp. 774–795, 2002.

[6] T. Kalis, K. Vannest, and R. Parker, "Praise counts: Using self-monitoring to increase effective teaching practices," *Preventing School Failure*, vol. 51, no. 3, pp. 20–27, 2007.

[7] L. Jenkins, M. Floress, and W. Reinke, "Rates and types of teacher praise: A review and future directions," *Psychology in the Schools*, vol. 52, no. 5, pp. 463–476, 2015.

[8] J. Morton, M. Mikolajczak, and O. Luminet, "New perspectives on the praise literature: towards a conceptual model of compliment," *Current Psychology*, 2020.

[9] S. Okada, Y. Ohtake, Y. Nakano, Y. Hayashi, H. Huang, Y. Takase, and K. Nitta, "Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*, 2016, pp. 169–176.

[10] Z. Tan, A. Goel, T. Nguyen, and D. Ong, "A multimodal lstm for predicting listener empathic responses over time," in *14th IEEE International Conferenceon Automatic Face and Gesture Recognition (FG '19)*, 2019, pp. 1–4.

[11] V. Perez Rosas, R. Mihalcea, and L. P. Morency, "Multimodal sentiment analysis of spanish online videos," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 38–45, 2013.

[12] M. Soleymani, K. Stefanov, H. Kang, J. Ondras, and J. Gratch, "Multimodal analysis and estimation of intimate self-disclosure," in *Proceedings of the 21st ACM International Conference on Multimodal Interaction (ICMI '19)*, 2019, pp. 59–68.

[13] H. Brugman and A. Russel, "Annotating multimedia / multimodal resources with elan," in *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC '04)*, 2004, pp. 2065–2068.

[14] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability," *Psychological bulletin*, vol. 86, no. 2, pp. 420–428, 1979.

[15] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT '19)*, 2019, pp. 4171–4186.

[16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] D. Bergstra, J.and Yamins and D. Cox, "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms," in *Proceedings of the 12th Python in Science Conferences (SciPy '13)*, 2013, pp. 13–20.