

WHAT IS THIS POSTER ABOUT?

This poster is about...

- 1) Analysis on contents of experts' feedback in Social Skills Training
- 2) Automated score estimation of social skills with multimodal features.

SOCIAL SKILLS TRAINING: SST

SST is

- A rehabilitation program to help people struggling with social skills
- More than 40 years history
- Improve social skills by role-playing specific situation followed by experienced trainer
- We focused on following basic skills (defined by Bellack et al. [1])
 - > LISTEN : listening to others
 - > TELL : expressing positive feelings
 - > ASK : requesting something to others
 - > DECLINE : declining requests
- Each social skills includes wide variety of techniques:
 - > Gaze activities
 - > Gestures
 - > Voice intonation
 - > Speech contents
 - > and so on...
- Limited number of trainers: inaccessible for everyone at anytime
 - > Long training is required to become SST trainer
 - > e.g.) One program lasts 8 to 12 weeks to complete
- Automated SST system has been studied for many years



Fig.1. Flowchart of SST

(Red part is a target of this research, modified from Fig. 1 of [2])

METHOD

OVERVIEW

Feedback generation: a critical but difficult process

- Wide and sparse variety which hampers machine learning performance

Our approach:

Step1. Find feedback tendency: An analysis of types of feedback contents

Step2. Estimate skill score: 7-component score estimation with multimodal features

(Step3. Generate feedback based on the scores, future work)

TECHNICAL DETAIL

Dataset: Human-human SST dataset

- 50 participants (16 Autism, 15 Schizophrenia, 19 Control, based on DSM-5)
- 7-component score annotated by 2 experienced SST trainers (Cohen's kappa: 0.84)
- #sessions: LISTEN-46, TELL-34, FAVOR-46, REFUSE-45

Step1. Distribution analysis of feedback content types:

- Counted the feedback types by checking every feedback transcription
- Our SST took summary feedback strategy (never be double-counted within a session)

Step2. Social skill score estimation:

- Model : Random forest for each component in each task (28 independent models)
- Input : 55 automated multimodal features inspired by result of step1.
- Output : 1 component score out of 7 components
- Misc.:

- > Leave-one-participant-out cross validation

> Feature selection based on trained model (n=5) <- typo correction of the paper

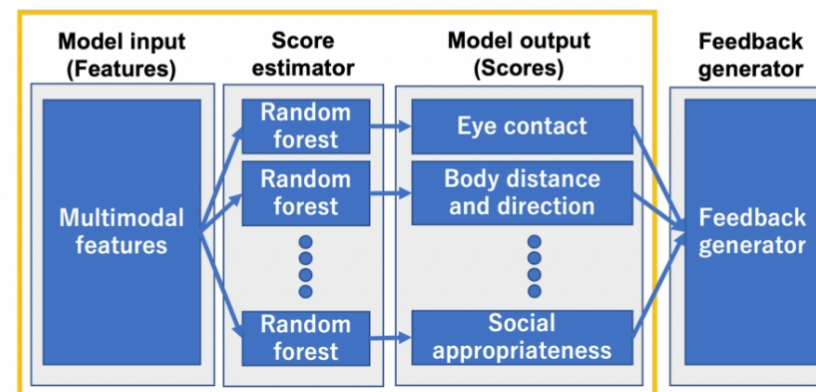


Fig.2. SST feedback system

("Feedback generator" is now in progress)

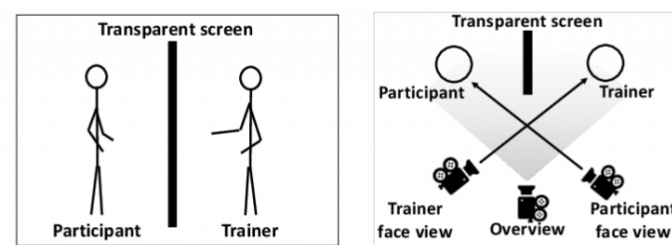


Fig.3. Schematic diagrams of SST dataset

(Left: side view, top view)

RESULTS

OVERVIEW

Step1. Distribution analysis of feedback content types

- Distribution was highly imbalanced

> Several frequently used feedback contents were observed

Step2. Social skill score estimation:

- Except for DECLINE, at least two models showed significantly correlated predictions (DECLINE is known as the more difficult task than others)

> Facial expression model in TELL task showed the highest correlation 0.53 and 0.26 R2 score

- Confirmed strong correlations -0.81 between RMSE and correlation coefficients after normalizing RMSE with the standard deviation of ground-truth values

Tab.1. (Step1.) Frequency count of feedback content types
(Freq.: frequency of feedback, Pos.: positive feedback, Neg.:)

Feedback types	Freq.	Pos.	Neg.
Appropriate facial expression	72	49	23
Seems contrite	35	21	14
Concrete speech content	34	21	14
Eye contact	32	30	2
Backchannel	25	17	8
Clearly refused	23	21	2
Checked whether they have time to talk	19	16	3
Gesture	16	13	3
Voice amplitude	5	5	0
Express gratitude	4	4	0
Nodding	4	4	0

Tab.2. (Step2.) Estimation results

(CORREL indicates Pearson's correlation coefficient, bold number indicate significant correlation p<0.05)

TASK	LABEL	R2	RMSE	CORREL	TASK	LABEL	R2	RMSE	CORREL
TELL	Eye contact	-0.20	1.07	0.06	ASK	Eye contact	-0.87	1.74	-0.22
	Body direct. and dist.	-0.18	1.12	0.16		Body direct. and dist.	-0.15	0.76	0.20
	Facial expression	0.26	1.17	0.53		Facial expression	0.09	1.41	0.42
	Voice variation	-0.05	1.40	0.24		Voice variation	-0.03	1.60	0.28
	Clarity	-0.41	2.25	-0.14		Clarity	0.03	1.55	0.37
	Fluency	-0.14	1.81	0.19		Fluency	-0.42	1.92	-0.06
LISTEN	Social appropriateness	0.14	1.39	0.40	DECLINE	Social appropriateness	0.18	1.25	0.47
	Eye contact	0.16	0.71	0.46		Eye contact	-0.11	1.34	0.28
	Body direct. and dist.	-0.15	0.91	0.17		Body direct. and dist.	-0.09	1.03	0.26
	Facial expression	-0.04	1.40	0.23		Facial expression	-0.19	1.76	0.14
	Voice variation	-0.18	1.70	0.14		Voice variation	-0.24	2.40	0.09
	Clarity	-0.10	1.44	0.14		Clarity	-0.20	2.26	0.08
	Fluency	-0.20	1.68	0.05		Fluency	-0.17	2.01	0.01
	Social appropriateness	0.13	1.11	0.40		Social appropriateness	-0.05	1.82	0.29

FUTURE WORK

What we've never solved yet

1. Need to consider more interactive features i.e. mutual gaze, voice amp. synchronization
2. Analysis between 7-component scores and actual feedback types

REFERENCES

- [1] Bellack et al. Social Skills Training for Schizophrenia: A Step-by-Step Guide, Guilford Press, 2004, 2nd edition
- [2] Saga et al. Multimodal Prediction of Social Responsiveness Score with BERT-based Text Features in IEICE Trans. Information and Systems, Human Communication, Mar. 2022, Vol.E105-D, No.3, pp.578-586