

HCG Symposium 2020

A-7-2

マルチモーダル情報を用いた ソーシャルスキルの客観的推定

*佐賀健志, *田中宏季, **岩坂英巳, *中村哲

*奈良先端科学技術大学院大学, **奈良県立医科大学精神医学講座



研究背景—ソーシャルスキルトレーニング

ソーシャルスキル：社会で生きていくための必須能力

- 1) うれしい気持ちを伝える
- 2) 頼み事をする
- 3) 相手の言うことに耳を傾ける
- 4) 不愉快な気持ちを伝える

ソーシャルスキルトレーニング (SST)

-> 目標：適切なソーシャルスキルの獲得

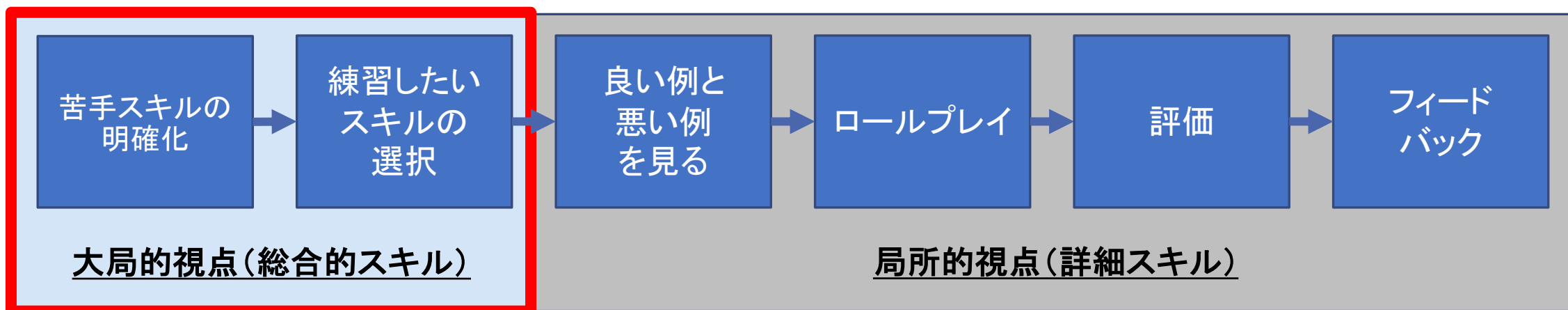
問題点：SSTセラピストが不足 -> アクセシ性が低い

解決策：SSTの自動化



研究目的—SST自動化にむけて

標準化された評価指標+機械学習でユーザーの得意不得意スキルを推定



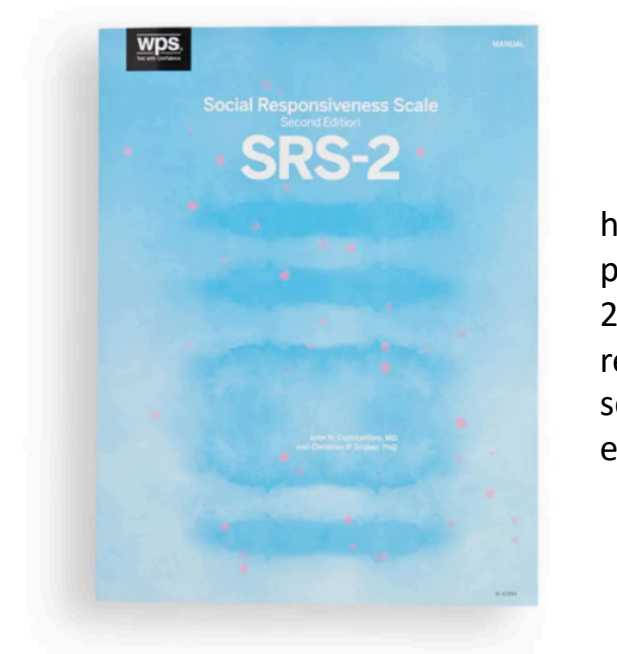
得意・不得意スキルの評価・判断

臨床現場 : 対話を通じたコミュニケーション

自動化システム : 標準化された評価指標 + 機械学習

対人応答性尺度 (SRS-2) [3]

- > 65問で構成される質問紙(標準化済み)
- > 自閉症傾向の重症度に関する評価尺度(健常者対応可)
- > 治療下位尺度も算出可能
 - 1) 社会的気づき (8 項目)
 - 2) 社会的認知 (12 項目)
 - 3) **社会的コミュニケーション** (22 項目)
 - 4) 社会的動機付け (11 項目)
 - 5) 興味の限局と反復行動 (12 項目)
- > すべての治療下位尺度を足し合わせることで**総合スコア**を得る
(自閉傾向が低いほどSRSスコアは低くなる。低いほど良い。)



<https://www.wpspublish.com/srs-2-social-responsiveness-scale-second-edition>

ベースライン

Tanakaら (2017) [1]

- 音声特徴量 : F0変動係数、声の大きさ、音声品質、ポーズ
- テキスト特徴量 : 語彙数、6文字以上の単語数、フィラーの数
- 画像特徴量 : 総フレーム数に対する笑顔比、頭の向き(ヨー、ピッチ)
- 機械学習モデル : 線形回帰
- 正解ラベル : セラピストの主観評価値

本研究のベースラインに設定するために変更
セラピストの主観評価値 → SRSスコア

提案手法

新規特徴量を導入

学習データ: 各1分間の発話動画(健康な被験者, 27人, 楽しかった出来事について)

入力特徴量: 28 (音声)^[4] + 5 (テキスト) + 22(画像) = 55

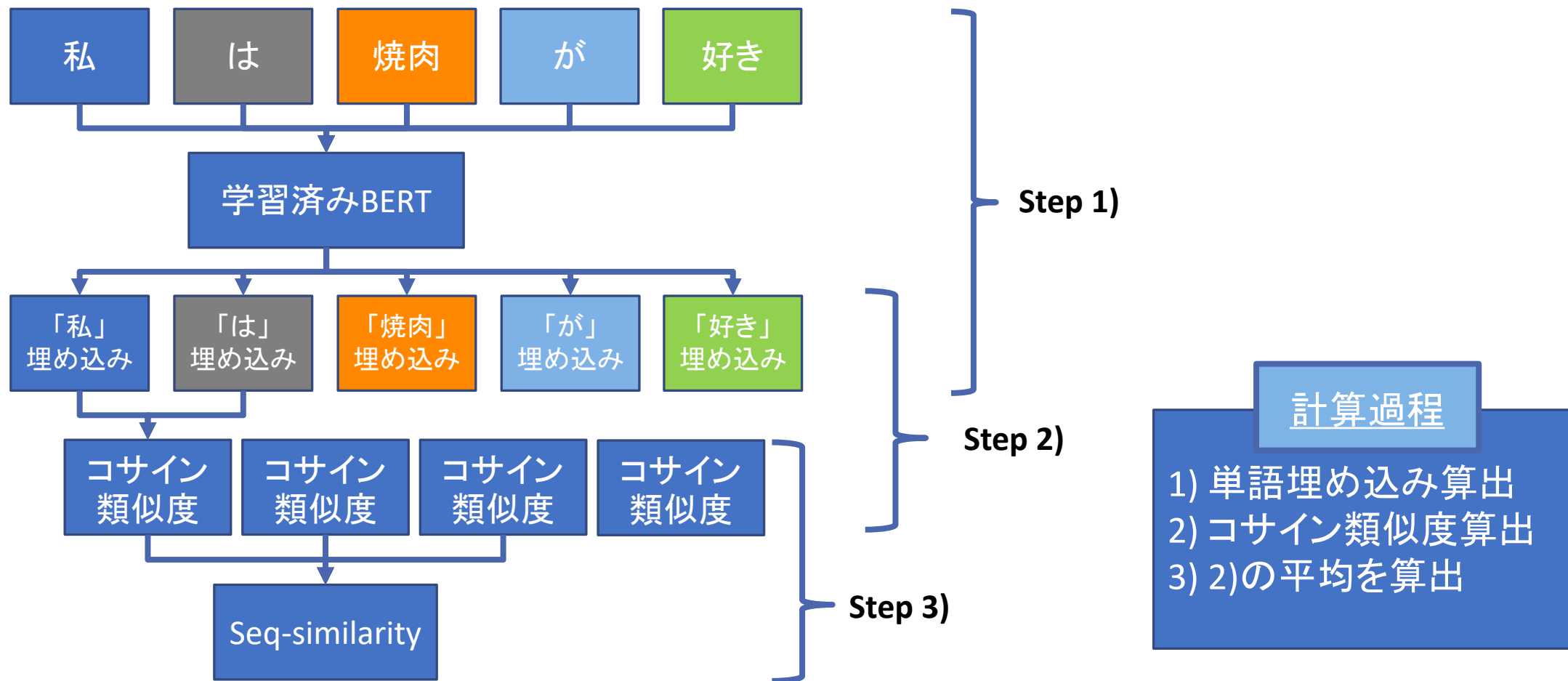
自動SSTシステムに新規導入した特徴量

- 面接評価システムで使用されていた音声特徴量の応用
- BERT埋め込みに基づく発話系列類似性 (Seq-similarity)

SST同様にマルチモーダル情報を使用

統合失調症などの精神疾患に顕著な「発散的思考」により脈絡なく話題が遷移^[5]
-> 発話内容の一貫性を評価

BERT埋め込みに基づく発話系列類似性



データセット [6]

- 対象** : 学生
- 被験者数** : 27人(男21、女6)
- タスク** : 「最近あった楽しかった出来事」について1分間話す
- 聞き手** : バーチャルエージェント(インタラクションなし)
- 書き起こし** : 手動
- 年齢** : 平均25.1、標準偏差2.13
- SRS(男)** : 平均65.9、標準偏差19.8
- SRS(女)** : 平均61.5、標準偏差21.0

重要スキル「うれしい気持ちを伝える」に相当

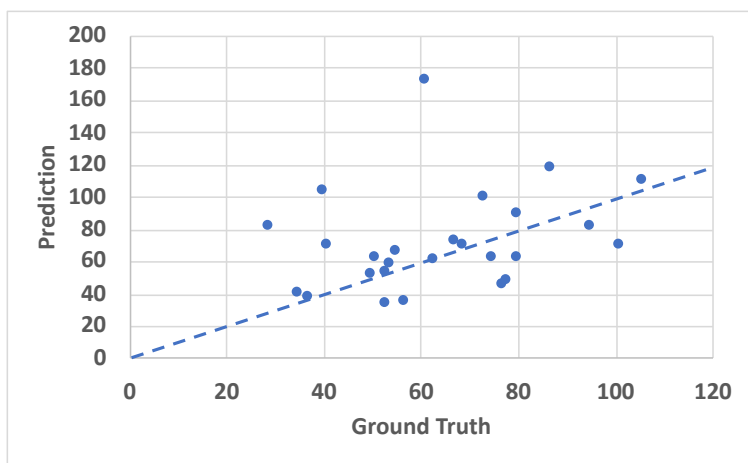
評価指標	主観評価値との相関
社会的気づき	-0.40
社会的認知	0.00
社会的コミュニケーション	-0.39
社会的動機づけ	-0.42
興味の限局と反復行動	0.10
総合スコア	-0.33

奈良先端科学技術大学院大学の倫理審査で承認

実験

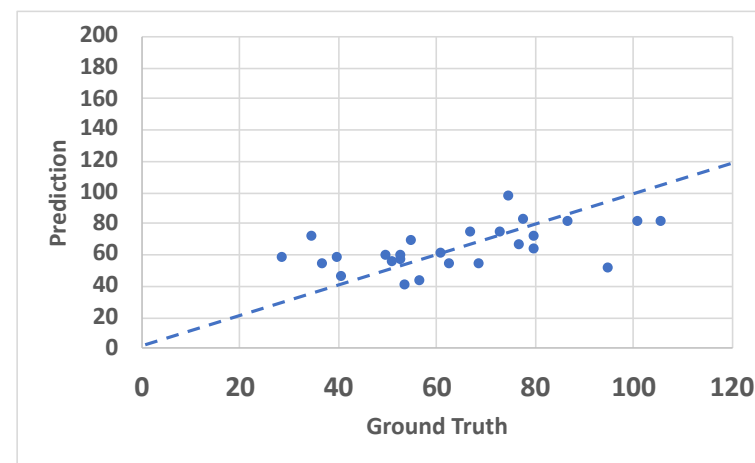
総合スコア予測

ベースライン [1]
(Tanakaら. 2017)



RMSE標準偏差	14.38
平均RMSE	27.34
Pearson相関係数	0.35
p値	0.08

提案手法



RMSE標準偏差	10.71
平均RMSE	17.41
Pearson相関係数	0.52
p値	0.01

いずれの指標でも性能向上が確認

考察

Seq-similarity, Energy, F2/F1 Mean — 重要特徴量の可能性

総合スコア予測 (モデル係数, 上位20項目)

Name	Coefficients
AU28	-409.6
A Seq-similarity	-115.6
Pitch	45.9
AU25	-42.7
AU15	-39.0
AU01	35.5
AU26	-33.6
AU45	31.9
AU17	-29.3
AU09	27.3

Yaw	-25.7
AU02	20.8
Smile Ratio	-12.7
B Energy	-12.4
AU07	12.2
AU12	-12.2
C F2/F1 Mean	8.0
AU05	6.1
AU06+12	-6.0
AU04	4.9

SRS : 高い→低い
ソーシャルスキル : 低い→高い

解釈

- A. 意味的一貫性を保って話す
- B. 大きい声で話す
- C. 長母音を多く使う [7]

予稿から
修正点

	Seq-similarity
総合スコア	-0.31
社会的コミュニケーション	-0.42

まとめと今後の展望

まとめ

-> マルチモーダル特徴量を用いてSRSスコアを相関係数0.52で予測

-> BERTに基づく発話系列類似性が重要な特徴量を捉えている可能性が示唆された

改善が必要なポイント

- 1) 精神疾患患者を対象とした検証
- 2) 結果に対する各特徴量の重要度および解釈についての検討

参考文献

- [1] Hiroki Tanaka et al., 2017, “Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders”, PLOS ONE 12, 8 (08 2017), 1–15. <https://doi.org/10.1371/journal.pone.0182151>
- [2] Mohammad Ali et al., 2018, “A Virtual Conversational Agent for Teens with Autism: Experimental Results and Design Lessons”, ArXiv, 1811.03046
- [3] John N. Constantino et al., 2012, “Social Responsiveness Scale, Second Edition (SRS-2)”, Back. Western Psychological Services.
- [4] IftexharNaim et al., 2015, “Automated Analysis and Prediction of Job Interview Performance”, IEEE Transactions on Affective Computing PP(042015). <https://doi.org/10.1109/TAFFC.2016.2614299>
- [5] Nemoto T, Mizuno M, Kashima H, “Qualitative evaluation of divergent thinking in patients with schizophrenia”, Behav Neurol. 2005;16(4):217-24. doi: 10.1155/2005/386932. PMID: 16518012; PMCID: PMC5478849.
- [6] Hiroki Tanaka, Hidemi Iwasaka, Hideki Negoro, Satoshi Nakamura, “Analysis of Conversational Listening Skills toward Agent-based Social Skills Training”, Journal on Multimodal User Interfaces, volume 14, issue 1, pp.73-82, 2020
- [7] T. Kagomiya, “Articulatory positions of japanese vowels as a function of duration computed from a large-scale spontaneous speech corpus,” 2015.