Automated Quality Assessment of CBT Sessions
through Highly Contextualized Language Representations

Nikolaos (Nikos) Flemotomos

University of Southern California
Signal Analysis & Interpretation Laboratory

March 17, 2022

Data Science for Mental Health Interest Group
@ The Alan Turing Institute

Why do we need to evaluate psychotherapy?

- lifetime prevalence of diagnosable mental disorders: more than 50%
- about 1 in 7 adults receives mental health services annually

Need for quality assurance

- more effective training
- more efficient supervision
- more positive clinical outcomes

image by cottonbro from Pexels

Why do we need to evaluate psychotherapy?

- lifetime prevalence of diagnosable mental disorders: more than 50%
- about 1 in 7 adults receives mental health services annually

Need for quality assurance

- more effective training
- more efficient supervision
- more positive clinical outcomes

- essential for improved performance: feedback to the therapist
  1. client progress monitoring
  2. performance-based feedback

image by cottonbro from Pexels

Behavioral coding

- psychotherapy: intervention based on spoken language
  ⟹ quality encoded in therapists' and patients' speech/language characteristics

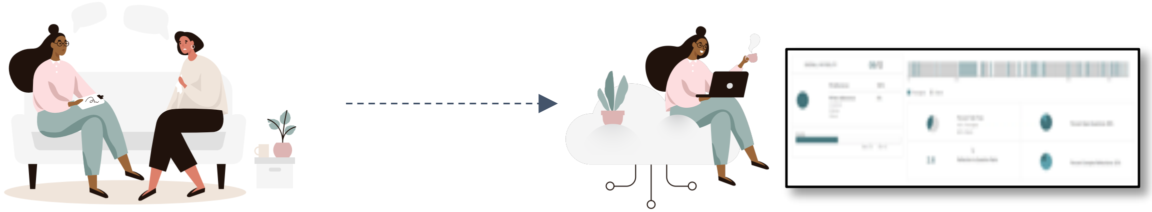
- quality assessment traditionally addressed by human raters using recorded sessions
  - time-consuming
  - cost-prohibitive

Behavioral coding

- psychotherapy: intervention based on spoken language
  ⟹ quality encoded in therapists' and patients' speech/language characteristics

- quality assessment traditionally addressed by human raters using recorded sessions
  - time-consuming
  - cost-prohibitive

⟹ *computational methods for automatic evaluation*



Behavioral coding in cognitive behavioral therapy

- CBT: one of the most popular psychotherapeutic approaches
- Aims at shifting the patient’s patterns of thinking

Monitoring CBT quality: Cognitive Therapy Rating Scale (CTRS)

- 11 session-level codes scored on a 7-point Likert scale (0=poor, 6=excellent)

abbreviation	meaning	
ag	agenda	*management and structure*
fb	feedback	
pt	pacing and efficient use of time	
hw	homework	
un	understanding	*good relationship*
ip	interpersonal effectiveness	
co	collaboration	
gd	guided discovery	*conceptualization*
cb	focusing on key cognitions and behaviors	
sc	strategy for change	
at	application of cognitive-behavioral techniques	

- $\sum_{i=1}^{11} \text{code}_i \geq 40 \implies$ competent delivery of CBT

Motivation

Existing methods...

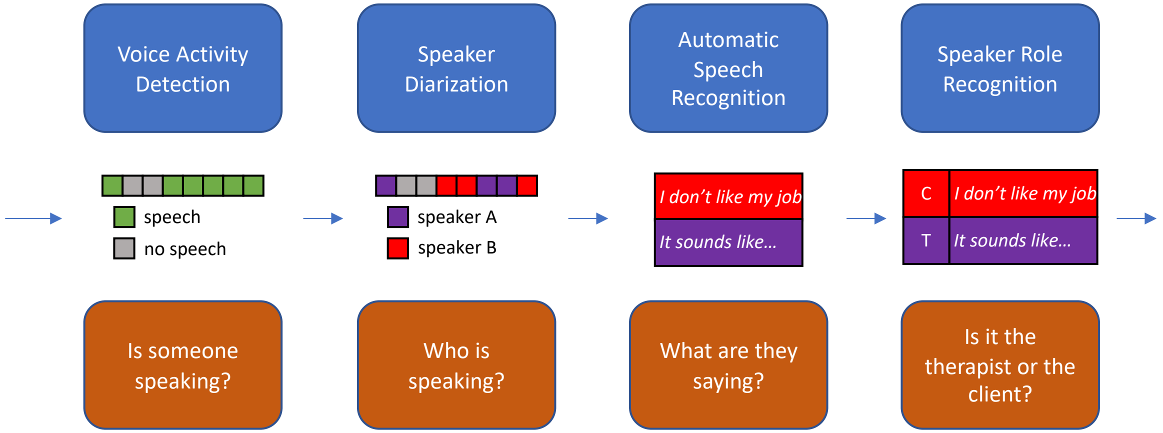
- use hand-crafted and/or sparse indicator features
  - can we better use context?
- model behavioral codes (and total CTRS) independently
  - but total CTRS in the sum of 11 codes!
- study CBT-related constructs appearing in short text excerpts
  - but a typical CBT session consists of hundreds of talk turns!

Rich transcription pipeline

- Our algorithms for automatic behavior coding are based on linguistic information (text).
- How do we get text from audio recordings?

Rich transcription pipeline

- Our algorithms for automatic behavior coding are based on linguistic information (text).
- How do we get text from audio recordings?



Voice Activity Detection → Speaker Diarization → Automatic Speech Recognition → Speaker Role Recognition

🟩 speech
⬜ no speech

🟪 speaker A
🟥 speaker B

I don't like my job
It sounds like…

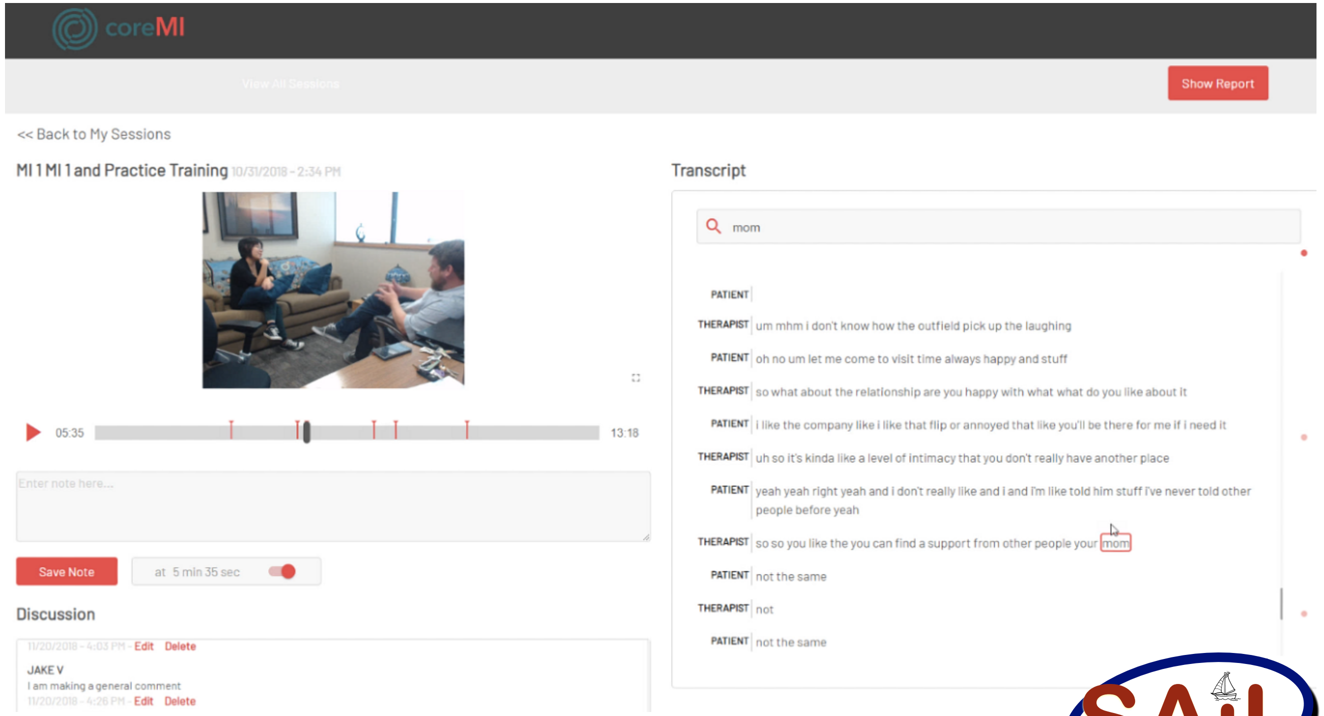
C *I don't like my job*
T *It sounds like…*

Is someone speaking?

Who is speaking?

What are they saying?

Is it the therapist or the client?



MI 1 MI 1 and Practice Training 10/31/2018 - 2:34 PM

Enter note here...

at 5 min 35 sec

Discussion

11/20/2018 - 4:03 PM - Edit Delete

JAKE V
I am making a general comment
11/20/2018 - 4:26 PM - Edit Delete

Transcript

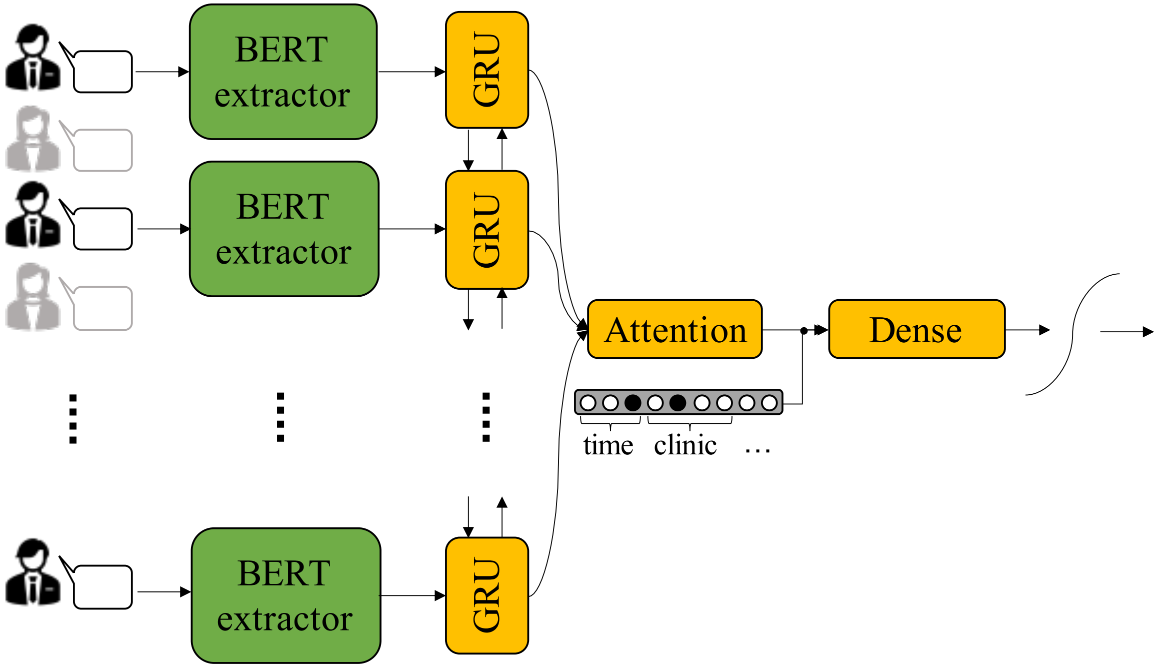
mom

PATIENT
THERAPIST um mhm i don't know how the outfield pick up the laughing
PATIENT oh no um let me come to visit time always happy and stuff
THERAPIST so what about the relationship are you happy with what what do you like about it
PATIENT i like the company like i like that flip or annoyed that like you'll be there for me if i need it
THERAPIST uh so it's kinda like a level of intimacy that you don't really have another place
PATIENT yeah yeah right yeah and i don't really like and i and i'm like told him stuff i've never told other people before yeah
THERAPIST so so you like the you can find a support from other people your mom
PATIENT not the same
THERAPIST not
PATIENT not the same

N. Flemotomos, V.R. Martinez, Z. Chen, K. Singla, V. Ardulov, R. Peri, D. Caperton, J. Gibson, M.J. Tanana, P. Georgiou, J. Van Epps, S.P. Lord, T. Hirsch, Z.E. Imel, D.C. Atkins, and S. Narayanan (2021). **Automated Evaluation of Psychotherapy Skills Using Speech and Language Technologies**, *Behavior Research Methods*

Model: single-task approach

- Directly model total CTRS as the binarized output variable.
- loss function: binary cross-entropy

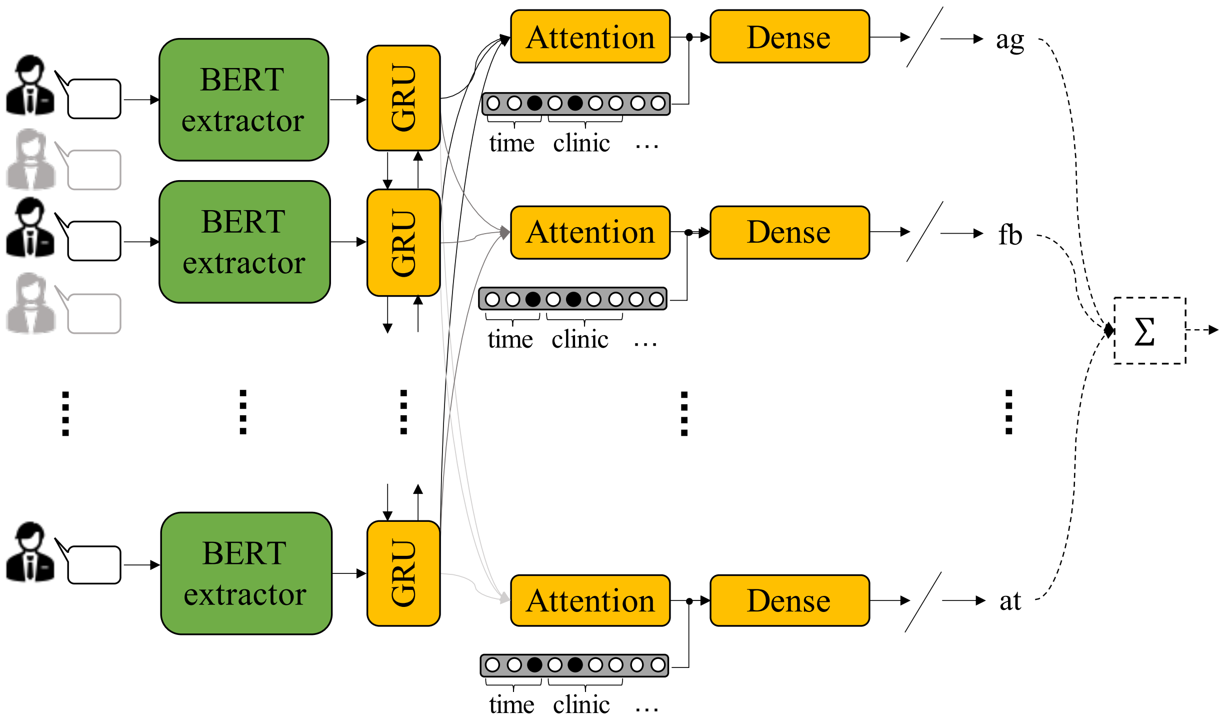


- BERT is adapted by continuing training on in-domain data (automatically transcribed psychotherapy sessions).

N. Flemotomos, V.R. Martinez, Z. Chen, T.A. Creed, D.C. Atkins, and S. Narayanan (2021).
Automated Quality Assessment of Cognitive Behavioral Therapy Sessions through Highly Contextualized Language Representations, *PLOS ONE*

Model: multi-task approach

- Model each CTRS code in a regression setting.
- Total CTRS is calculated as the (unweighted) sum and then binarized.
- loss functions: mean squared error

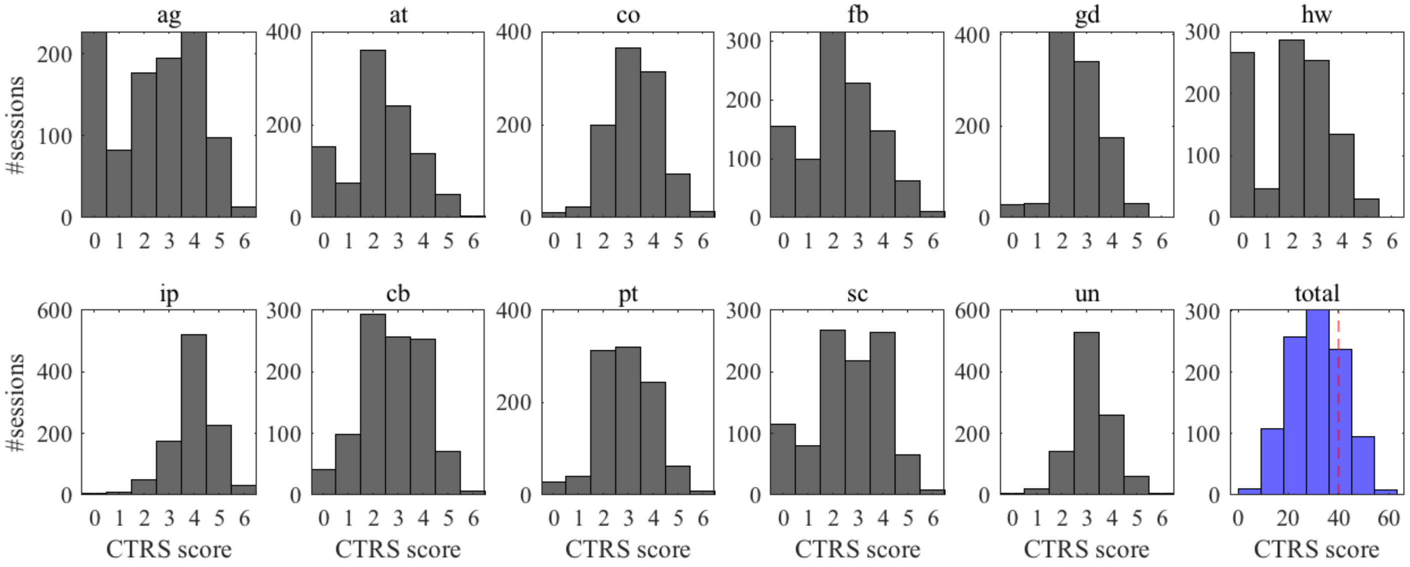


- advantage: higher interpretability

N. Flemotomos, V.R. Martinez, Z. Chen, T.A. Creed, D.C. Atkins, and S. Narayanan (2021).
Automated Quality Assessment of Cognitive Behavioral Therapy Sessions through Highly Contextualized Language Representations, *PLOS ONE*

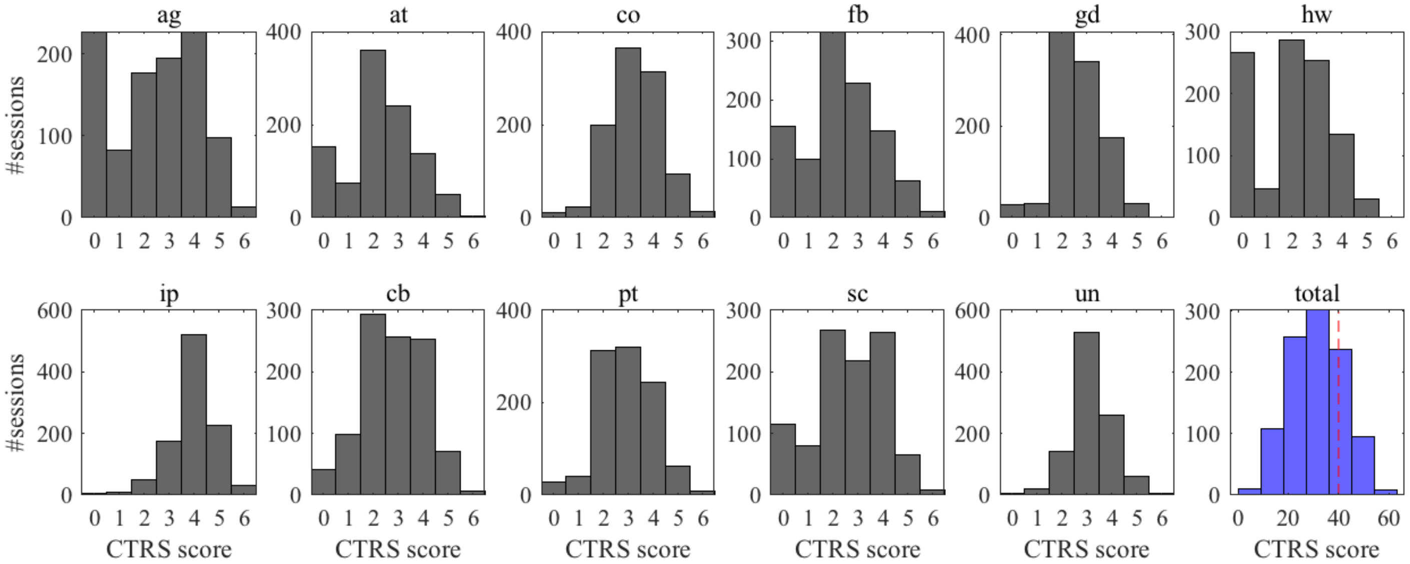
CBT dataset

- 1,018 recorded, manually coded CBT sessions (mean dur = 41.5min), automatically transcribed
- available metadata
  - *clinic:* 383 therapists across 25 clinics
  - *level of care:* 6 categories (inpatient, outpatient, school-based, etc.)
  - *population:* 9 population groups (child, adult, substance use, etc.)
  - *assessment time wrt CBT training:* 7 timestamps (pre-workshop, post-workshop, 1 month after, etc.)



CBT dataset

- 1,018 recorded, manually coded CBT sessions (mean dur = 41.5min), automatically transcribed
- available metadata
  - *clinic:* 383 therapists across 25 clinics
  - *level of care:* 6 categories (inpatient, outpatient, school-based, etc.)
  - *population:* 9 population groups (child, adult, substance use, etc.)
  - *assessment time wrt CBT training:* 7 timestamps (pre-workshop, post-workshop, 1 month after, etc.)



- 100 additional CBT sessions used to adapt the ASR pipeline
- 4,263 recorded, non-coded psychotherapy (not necessarily CBT) sessions for BERT adaptation

Results – I

		all utterances		therapist-only utterances	
utterance representation	metadata info	single-task	multi-task	single-task	multi-task
BERT-base	✗	63.43	61.03	63.88	62.40
	✓	65.42	70.13*	66.80$^\#$	71.25*
adapted BERT	✗	64.10	62.04	65.52	63.76
	✓	66.94$^\#$	71.56*	68.52*	**72.61***

F$_1$ score (%) – 10-fold cross validation. $^\#$p$<$0.05, *p$<$0.01

Results – II

proposed technique	no	yes	relative improvement
adapt BERT	65.54	66.88	+2.04%
metadata info	63.27	69.15	+9.29%
multi-task	65.58	66.85	+1.94%
only therapist	65.58	66.84	+1.92%

each row: mean F_1 score (%) across all the remaining 2^3=8 combinations when the corresponding technique is or is not applied

Results – II

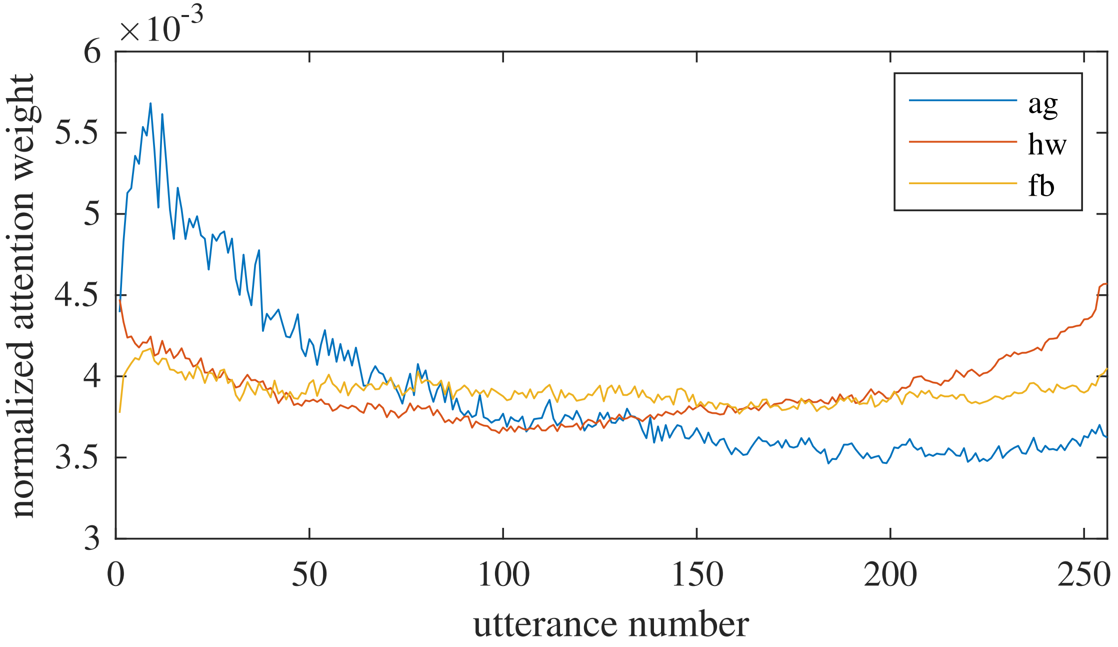
proposed technique	no	yes	relative improvement
adapt BERT	65.54	66.88	+2.04%
metadata info	63.27	69.15	+9.29%
multi-task	65.58	66.85	+1.94%
only therapist	65.58	66.84	+1.92%

each row: mean F_1 score (%) across all the remaining 2^3=8 combinations when the corresponding technique is or is not applied

- adapted BERT > pre-trained BERT-base
  - fine-tuned both on the domain *and* on ASR-induced errors
- therapist-only utterances > all utterances
  - CTRS codes are focused only on therapist behavior
- incorporation of metadata information beneficial
  - however, such information may not be available in general
- multi-task > single-task *when* metadata is provided
  - metadata improve robustness when predicting each code ⟹ overall robustness

Localization of CTRS codes

- CBT is a highly structured psychotherapeutic approach
  ⟹ reflected in several of the CTRS codes
- Using the attention mechanisms, we can identify salient utterances
  ⟹ reveal this structure,
  ⟹ examine how the practitioner focuses on different aspects of CBT throughout therapy



Mean attention weights across all the sessions

Practical and ethical implications – I

- Is it acceptable to use patients' sensitive data?
  - all patients and therapists sign a consent form
  - approved by Institutional Review Board (sufficient?)
  - all data are de-identified wrt patients

image from the Noun Project. creator: SBTS

Practical and ethical implications – I

- Is it acceptable to use patients' sensitive data?
  - all patients and therapists sign a consent form
  - approved by Institutional Review Board (sufficient?)
  - all data are de-identified wrt patients

- What if such a system is used to blindly evaluate a therapist? That could even mean loosing their job!
  - the goal is not to replace human supervision, but rather augment the supervisor's capabilities and offer a tool for self-assessment
  - users should be adequately trained to understand the meaning of automatically generated feedback and evaluation scores

image from the Noun Project. creator: SBTS

Practical and ethical implications – II

- How to mitigate potential biases?
  - adaptation to the actual use case
    (e.g., perceptions about psychotherapy differ across cultures)
  - employ large and diverse pools of human coders
  - fairness through unawareness (both for models and for annotators)

Practical and ethical implications – II

- How to mitigate potential biases?
  - adaptation to the actual use case
    (e.g., perceptions about psychotherapy differ across cultures)
  - employ large and diverse pools of human coders
  - fairness through unawareness (both for models and for annotators)

- Any additional requirements before using in clinical settings?
  - incorporate confidence metrics and quality safeguards of the model
  - users should be able to question model predictions (human-in-the-loop)

Conclusions

- Introduced a model for automatic evaluation of CBT sessions and compared various configurations
- Demonstrated the importance of context – both linguistic and non-linguistic through available metadata

Conclusions

- Introduced a model for automatic evaluation of CBT sessions and compared various configurations

- Demonstrated the importance of context – both linguistic and non-linguistic through available metadata

Future Vision

- Widespread adoption of psychotherapy evaluation systems in clinical practice, leading to improved quality of services

- under a proper ethical and practical framework, ensuring
  - data privacy
  - bias mitigation
  - prudent usage and interpretation
  - proper error handling

Thank you!