Automated Quality Assessment of Cognitive Behavioral Therapy Sessions Through Highly Contextualized Language Representations

Nikolaos Flemotomos[1], Victor Martinez[1], Zhuohao Chen[1], Torrey Creed[2], David Atkins[3], Shrikanth Narayanan[1]

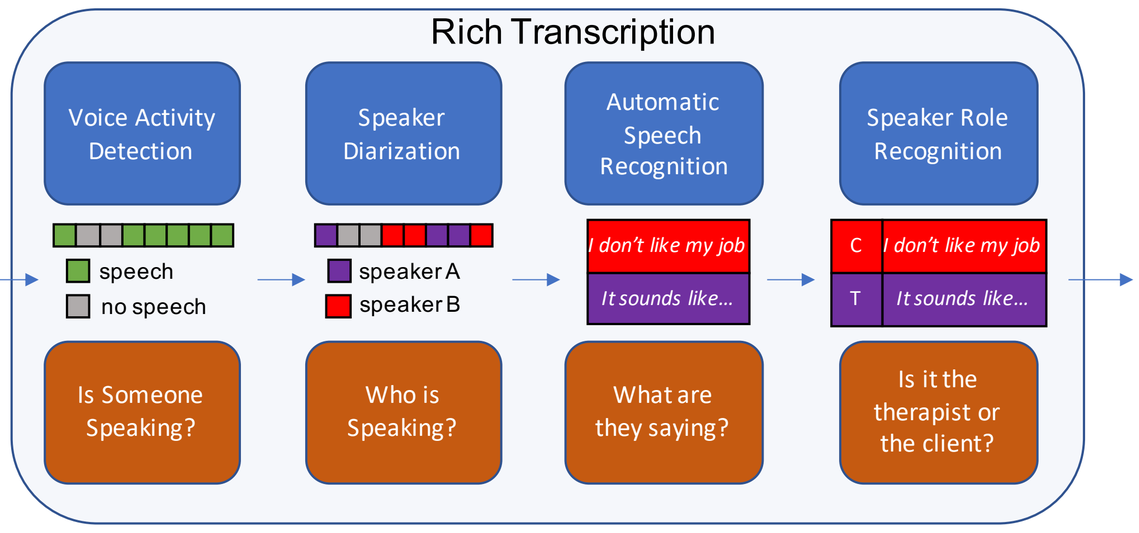
[1]Signal Analysis and Interpretation Lab (SAIL), University of Southern California
[2]Department of Psychiatry, University of Pennsylvania
[3]Department of Psychiatry and Behavioral Sciences, University of Washington

Psychotherapy Evaluation & CBT

- psychotherapy
  communication-based intervention to treat mental health disorders
- quality assessment
  - more effective training
  - more efficient supervision
  - more positive clinical outcomes
- Cognitive Behavior Therapy (CBT)
  one of the most popular psychotherapeutic approaches
- monitor CBT quality: Cognitive Therapy Rating Scale (CTRS)
  - 11 session-level codes (e.g., understanding, collaboration, homework, ...)
  - scored on a 7-point Likert scale
  - $\sum_{i=1}^{11} \text{code}_i \geq 40 \Rightarrow$ competent delivery of CBT [total CTRS]
  - *focus on this binary classification problem*
- existing methods
  - based on hand-crafted / sparse indicator features
  - model total CTRS independently of the individual codes
  - analyze short text excerpts - not actual therapy sessions
- proposed approach
  - use contextualized language models (BERT)
  - provide therapist- and session-related metadata
  - multi-task approach: model total CTRS as the sum of the 11 codes

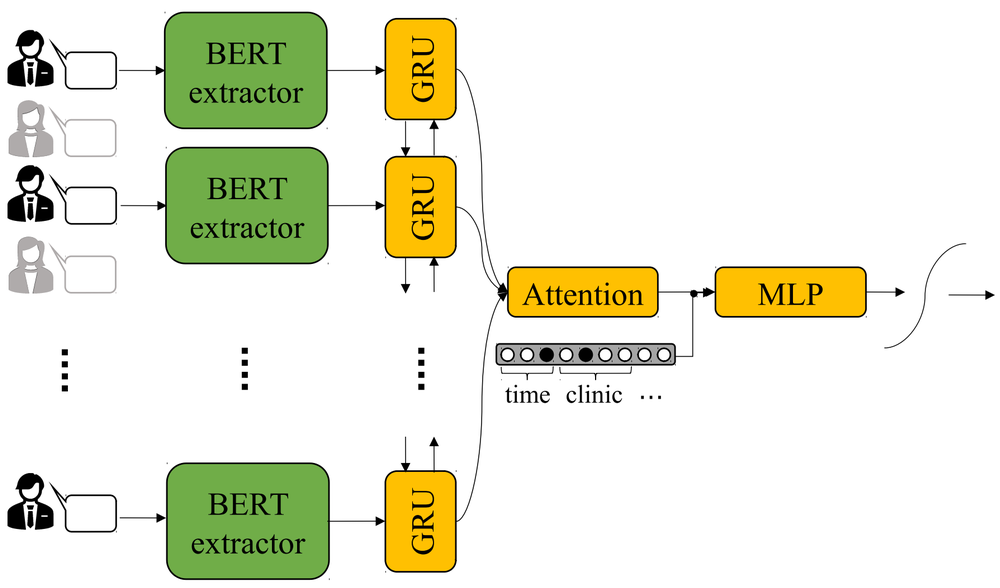
Datasets & Preprocessing

- 1,018 CBT sessions recorded in community health care centers, accompanied by CTRS scores
  largest-to-date study focusing on automated CBT evaluation
- 100 additional sessions to adapt the transcription pipeline
- 4,263 recorded, non-coded sessions (not necessarily CBT) to adapt the language model
- all sessions automatically transcriped via pipeline adapted to psychotherapy



Single-Task Approach

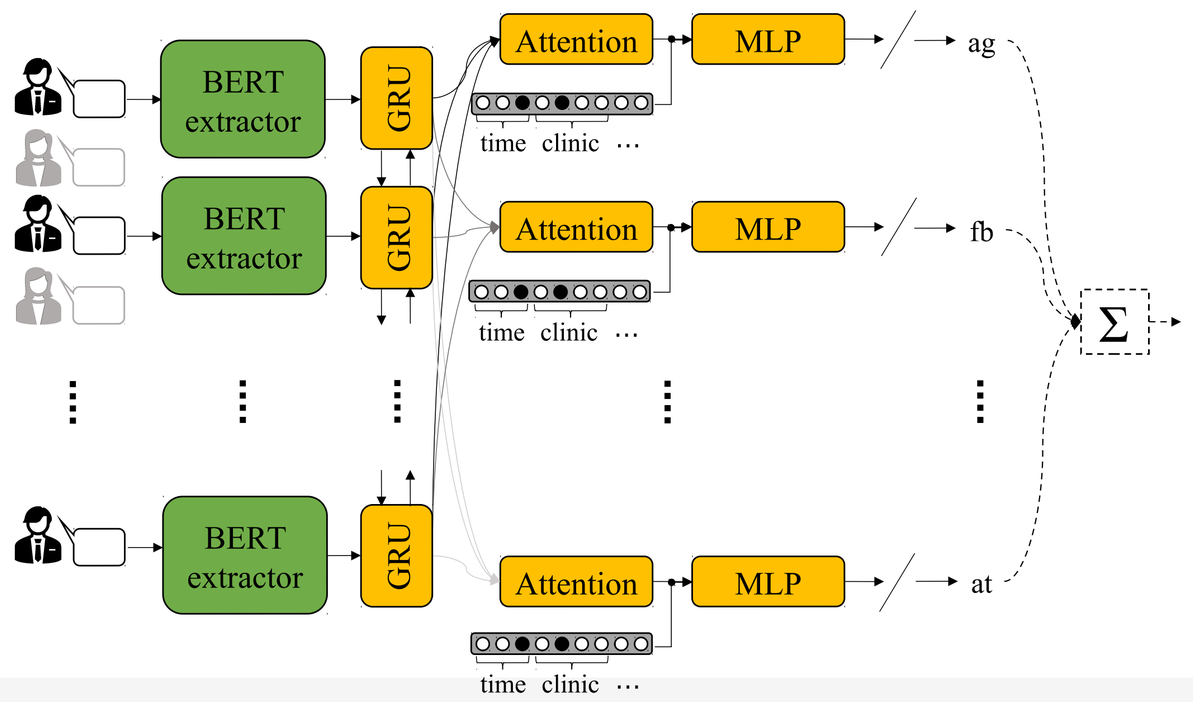
- directly model total CTRS as the binarized output variable



- BERT adapted on the domain and on ASR-induced errors

Multi-Task Approach

- total CTRS = (unweighted) sum; then binarize



Results

BERT model	metadata info	all utterances single	all utterances multi	therapist-only single	therapist-only multi
base	✗	63.43	61.03	63.88	62.40
	✓	65.42	70.13*	66.80†	71.25*
adapted	✗	64.10	62.04	65.52	63.76
	✓	66.94†	71.56*	68.52*	**72.61***

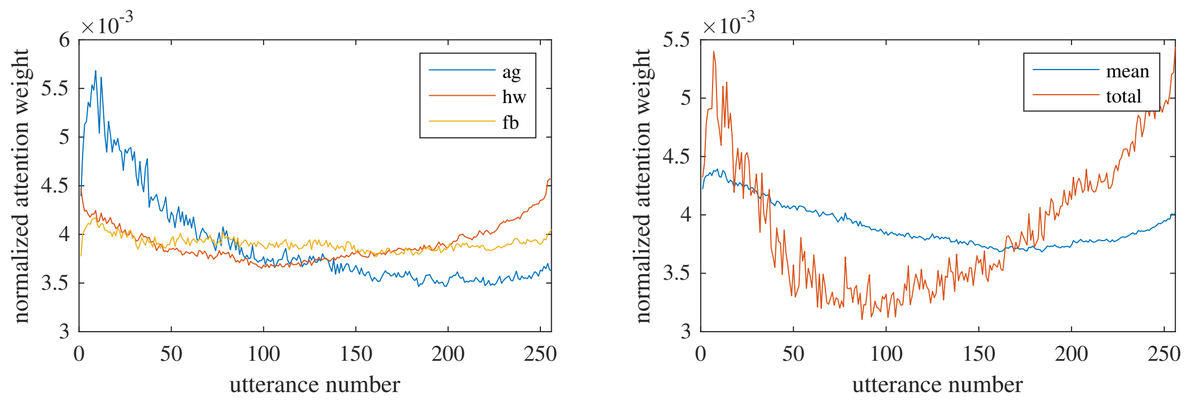
F_1 score (%) – 10-fold cross validation. $^{\dagger}p < 0.05$, $^{*}p < 0.01$

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 (*yes*) or is not (*no*) applied

Interpretable Models: CTRS Localization

- use attention mechanism to identify salient utterances
- reveal CBT structure



mean attention weights across all the sessions
(left) CTRS codes agenda, homework, feedback; (right) mean of the 11 codes, total CTRS

Future Vision

- widespread adoption of similar systems in clinical practice for
  - self-assessment
  - assisted supervision

  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

flemotom@usc.edu
October 29, 2021
doi: 10.1371/journal.pone.0258639