Language Features for End-to-End Evaluation of Cognitive Behavior Psychotherapy Sessions

Nikolaos Flemotomos¹, Victor R. Martinez², David C. Atkins³, Torrey A. Creed⁴, Shrikanth Narayanan¹,²
¹ Department of Electrical Engineering and ² Department of Computer Science, University of Southern California, Los Angeles, CA, USA
³ Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA
⁴ Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

Motivation & Background
• Cognitive Behavior Therapy (CBT) is a psychotherapy treatment that uses cognitive change strategies to address mental health problems¹. Quality assessment is traditionally addressed by human raters who evaluate recorded sessions along 11 behavioral codes, defined by the Cognitive Therapy Rating Scale (CTRS).
• We examine how linguistic features can be effectively used to develop an automatic competency rating tool for CBT.
• Experiments are conducted on manual transcripts and on automatically derived ones, thus introducing an end-to-end approach.

Method

Features
• Unigrams with tf-idf weighting²
• GloVe embeddings³
• Linguistic Inquiry and Word Count (liwc) features⁴
• Psycholinguistic Norm Features (pnf)⁵
• Counts of Dialogue Acts (da)⁶
• Counts of non-lexical (cnt-nlex) tags in transcripts

Datasets & Preprocessing
• adout set: 386 adult-outpatient sessions from 131 therapists
• trans set: manually transcribed subset of adout
92 sessions from 70 therapists

Results

Conclusions
• tf-idf_T almost always yields the best results among the independent features, but NOT for the highly skewed codes
• Best performance usually achieved through a fusion method
• For adout combining T’s and C’s features seems beneficial, but not for trans. Diarization and role matching errors?

Future efforts will focus on
• regression (‘how good is a session?’)
• interpretability of the results: feedback for improvement

References
¹ A. T. Beck, Cognitive Therapy of Depression, Guilford press, 1979
² G. Salton et al., Introduction to Modern Information Retrieval, McGraw-Hill Inc., 1986
³ J. Pennington et al., “GloVe: Global vectors for word representation,” in Proc. EMNLP, 2014

Results for the fusion techniques are reported only if better than the best independent classifier.

Features

Datasets & Preprocessing
• CTRS codes are manually scored on a 7-point Likert scale (0-6)⁷
• We binarize them labeling as negative (‘bad’) sessions with a threshold of 40.

Future efforts will focus on
• regression (‘how good is a session?’)
• interpretability of the results: feedback for improvement

References
¹ A. T. Beck, Cognitive Therapy of Depression, Guilford press, 1979
² G. Salton et al., Introduction to Modern Information Retrieval, McGraw-Hill Inc., 1986
³ J. Pennington et al., “GloVe: Global vectors for word representation,” in Proc. EMNLP, 2014

Method

Features

Datasets & Preprocessing
• CTRS codes are manually scored on a 7-point Likert scale (0-6)⁷
• We binarize them labeling as negative (‘bad’) sessions with a threshold of 40.

Future efforts will focus on
• regression (‘how good is a session?’)
• interpretability of the results: feedback for improvement

References
¹ A. T. Beck, Cognitive Therapy of Depression, Guilford press, 1979
² G. Salton et al., Introduction to Modern Information Retrieval, McGraw-Hill Inc., 1986
³ J. Pennington et al., “GloVe: Global vectors for word representation,” in Proc. EMNLP, 2014

Results for the fusion techniques are reported only if better than the best independent classifier.