Towards Semi-Automatic Monitoring of Delivery Behavioral Interventions
Carlos Gallo, Cady Berkel, Irwin Sandler, Hendricks Brown
Department of Psychiatry and Behavioral Sciences, Northwestern University / REACH Institute, Arizona State University
carlos.gallo@northwestern.edu

Introduction

Traditional methods for monitoring implementation of evidence-based programs require labor-intensive quality assessments. These assessments generally involve human observation, hence becoming a major bottleneck in the monitoring of implementation. We present the development of computer-based methods for measurement quality measurement.

Firstly, we describe 7 principles of automation necessary to systematically identify critical dimensions for analysis.

1. Be relevant to the program’s theory
2. Be burdensome for human coding
3. Be reliable across raters (>IRR .75)
4. Be a driver of the dimension
5. Vary across the sample
6. Be operationalizable
7. Predict program outcomes

Methods

Based on human-based ratings of a subset of 22 sessions of the New Beginning Program (NBP) for divorcing parents [2], we identified linguistic behavior that the group leader delivering the program should follow: a) Provide helpful examples, b) Indicate belief in parent’s ability. These dimensions met 7 Automation Principles, including having the highest inter-rater reliability and having predicted the relevant outcomes such as parent’s retention and positive engagement in home practice. We extracted linguistic features from this set of transcripts and used a Machine Learning KNN algorithm to classify utterances where quality indicators demonstrated high and low quality for the targeted dimensions.

Data

Creation of Corpus: Sessions of New Beginnings Program delivery was recorded, transcribed, and rated for quality by humans. From the transcripts, we selected the speech from the group leader, segmented into utterances, and applied the session-level quality rating to each utterance. A sample of the data is shown in the table below. In this example, there are two sessions transcripts, six utterances from group leader, and six quality ratings (3 high and 3 low).

Machine Learning Classifier

Training and Testing a machine classifier. There are a series of steps necessary to train and test a machine classifier. In general terms, we need to:
1. Create the Corpus
2. Preprocess of the Corpus - Stem words
3. Extract features for KNN Model [5]
4. Selecting the training & testing sets
5. Train the model
6. Test the model

Results and Discussion

Can machine learning recognize high versus low quality dimensions in session delivery transcripts?

Indicated Belief Dimension - We found that based on 22 transcripts that were coded for Indicated Belief quality dimension, the KNN (K-Nearest Neighbor) algorithm correctly classified 83% of the utterances to training documents. This results suggests that monitoring quality can be effectively with computational methods.

Future Work

We will apply these methods to other dimensions such as “Provide Helpful Examples.” Our goal is to be able to rate quality directly from the audio signal without the need of human transcription. First, we demonstrate that linguistic features can be measured automatically. Second, we need to adapt, test, and validate automatic transcription for measuring quality. Another area to explore is the interaction between group leaders and parents. Dialog Modeling can provide other measurable metrics that correlate with quality of delivery.

References