Machine learning and evidence-based counseling: Scaling up fidelity monitoring of psychosocial interventions via technology

Dave Atkins, PhD
datkins@uw.edu

U of Washington
Co-Director, BRiTE Center
https://www.brite.uw.edu

CEO, Lyssn.io
https://lyssn.io/

BRiTE CENTER

Graphic by Alice Gray
Acknowledgments and Disclosure

Grants/Funding:
- NIH / NIAAA R01 AA018673
- NIH / NIDA R34 DA034860
- NIH / NIAAA K02 AA023814
- UW Innovation Fund Award

Disclosure:
I am a co-founder of, and have an equity stake in, Lyssn.io, which is a technology start-up focused on developing and implementing tools to support evidence-based counseling.
Overview

1. The challenge of evaluating psychosocial interventions

2. Quantifying the raw information in human interactions

3. Technology for automated, performance-based feedback

4. Implementation of technology in community clinics to support training and QA
Behavioral health problems are massive and disabling

20% of Americans (60M) struggle with a mental health or addiction problem

Behavioral health problems cause over 28% of global burden of disease

We spend $8B-$10B per year on more than 100M counseling sessions.

Are they any good?
There are 100M counseling sessions each year, and we do not know the quality of any of them.

Empathy by Provider in Community Addiction Treatment

Future Directions for dissemination: “...objective assessment of fidelity including clinician competence including [who] completes training, achieve competence, and sustain competence.” McHugh & Barlow (2010)
Performance-based feedback for training and supervision

Quality assurance / improvement of service delivery

Predicting which treatments are likely to succeed and why

For payors, knowing what they are paying for

Traditional evaluation via human coding: Slow, expensive, not used in real-world
Natural Language Processing

Paralinguistics and vocal acoustics

Technological advances now enable automated processing of counseling sessions

Automatic Speech Recognition

Machine (‘deep’) Learning
What is the raw data – the basic building blocks – of psychotherapy?

**Words:**

“Wow, I can not imagine how difficult that must have been to lose your kids.”

**Tone:**

(paralinguistic information)

“Wow, I can **not** imagine how difficult that must have been to **lose** your kids.”

**Posture:**

(nonverbal information)

**Interaction:**

How are words, tone, and nonverbal behavior linked and dynamically evolving during the session?
How do we computationally quantify the raw data of psychotherapy?

**Words:** Individual words and short phrases (n-grams) + text-based machine learning

**Tone:** Speech signal processing methods for estimating “acoustic features”

**Posture:** Quantify movement and facial emotions from video (not current focus of our work)

**Interaction:** Synchrony, entrainment, and dynamic systems models for dyadic linkage and influence
Motivational interviewing is fundamentally linguistic and its fidelity is measured in its language.

Client: “I wouldn’t mind coming here for treatment, but I don’t want to go to one of those places where everyone sits around crying and complaining all day.”

Counselor: “You don’t want to do that... so, you’re kind of wondering what it would be like here.”

Client: “Yeah”
Text-based machine learning methods quantify ‘lexical features’ to be used in predictive models.

I wouldn’t mind coming here for treatment but I don’t want to go to one of those places where everyone sits around crying and complaining all day.

So you’re kind of wondering what it would be like here.

You don’t want to do that.

Yeah.

Who is speaking? What was previous code?

Directly overlapping words or parts of speech

Words in local context (before or after)

Individual vocabulary words and common 2 to 3 word phrases

Contextual n-gram features:

Counselor:Before:"YOU DO NOT WANT"
Counselor:Before:"YOU ARE"
Client:After:"YEAH"

n-gram features:

YOU WONDERING LIKE YOU ARE
KIND OF WONDERING IT WOULD BE LIKE

Contextual n-gram features:

Client:Before:"I WOULD NOT"
Client:Before:"I DO NOT WANT"
Client:After:"YEAH"

Counselor:Before:"YOU DO NOT WANT"

Meta features:

Counselor:Client
CL_RE:Client

Similarity features:

here ADVERB
Machine learning methods can be quite accurate at replicating human coding.

**Machine Learning Models**

1. **Latent Dirichlet allocation (LDA; aka, 'topic models')**
   - Atkins et al. (2014).
   - Implementation Science

2. **Conditional random fields (CRF) model**
   - Sequential decoding of 'states' (similar to hidden Markov models)
   - Can et al. (2016).
   - Journal of Counseling Psychology

3. **Recursive neural networks (RNN) model**
   - 'Deep learning' using multiple hidden layers
   - Tanana et al. (2016).
   - Journal of Substance Abuse Treatment

**Code Set**

- Facilitate
- Giving_Info
- Open_Question
- Closed_Question
- Reflection_Simple
- Affirm
- Reflection_Complex
- Confront

**Model Comparison**

- **CRF**
- **LDA**
- **RNN**

**Code Frequency**

- 0%: 2000
- 50%: 4000
- 100%: 6000

**Human Reliability**
Spoken language technologies can also inform psychotherapy mechanisms

Empathy involves:

1) affective synchrony
2) cognitive understanding

Lord et al., *J of Sub. Abuse Treat.* 2015

Language style matching

Mean Language Style Synchrony

Lord et al., *J of Counseling Psychology* 2013

Vocally encoded arousal
Immediate, performance-based feedback requires a coding and prediction pipeline

**Processing Steps**

**Signal Acquisition**

**Voice Activity Detection**

**Segmenting and Role Assignment**

**Automatic Speech Recognition**

**Predictive Models**

**CORE-MI Report**

---

**Goal of Each Step**

**Digital Recording to Server**

**Is Someone Speaking?**

**Who is the Speaker?**

**What did they say?**

**Use words to predict Codes**

**Feedback to Counselor**

---

**Data Collection**

**Speech Signal Processing**

**Prediction**

**Report Generation**

---

**Speech Signal Processing Tasks**

Not just prediction
Predicting empathy using a fully automated system – “Sound to Codes”

Using 200 sessions from John Baer et al. MI training study:  
\[ r = 0.65 \] (empathy)  
Xiao et al., 2015, *PLOS ONE*

Using 20 new, role-played sessions, with no models fit to sessions:  
\[ r = 0.61 \]

Using 1,384 sessions from 6 MI RCT or training studies:  
\[ r = 0.59 \]
Good news! You are Empathy = 4.2!

If this is ever going to get used in the real world, we need an effective report for the therapists...

CORE-MI Interactive Report:
1. Designed for computer or tablet
2. Overall fidelity statistics
3. Detailed session view, with talk-turn statistics

Let's take a look at a session from Steve Rollnick

http://sri.utah.edu/psychtest/misc/smoker1.html
Pilot feasibility study and user-experience feedback very positive

**Feasibility pilot:**
- 20 therapists
- Brief standardized patient session
- Automated feedback + interview

**Comparison of Expert and Novice Providers on Machine-Generated Codes**

**Preliminary user-experience data (% Agree):**
- “I found the report representative of my clinical performance”: 85% (17/20)
- “Overall, I was satisfied with the feedback”: 100% (19/19)
- “If tool was available, I would use it in my practice”: 89% (17/19)
- “[Fidelity measures] were easy to understand”: >80%*
Based on feedback, are there specific ways you see to improve your skills?

- “Talk less but use talk time more efficiently by reflecting more asking more open ended questions and complex reflect more.”
- “Empathy score was unclear as to what my score meant in general. Did I have a good level of empathy or a "bad" level?”

Anything you’d like added in the feedback?

- “Numerical markers of vocal synchrony and vocal intensity.”
- “If there was a way of identifying % of change talk on the part of the client that would be helpful.”
- “(1) I would love to have specific behavioral tags on the session view timeline! (2) Would love to have audio connected to session timeline.”
- “If there was a way to track the client’s “aha” moments and connections or any other significant client interactions to see what technique lead to the progress.”
Anything else you’d like to tell us?

- “I really like this kind of quantitative feedback as part of counselor training (or counselor eval in the case of working professionals). The facts tell an important part of the story—and along side personal feedback in the case of students I think this offers some insights that cannot be compared to the qualitative feedback of a supervisor. I also like the speed with which I got this info. Such quick turn around means that I can recall much of the session for which I am being evaluated making the feedback more impactful and improvements more likely.”
- “THIS IS GOING TO BE AWESOME!!!!!”
- “It’s hard to argue with a computer…”
Current work focuses on implementing and evaluating technology in community clinics

**Pilot projects for training and supervision**

MI (and other EBPs) for opiate addiction at publicly-funded community clinic

General mental health and addiction services + training clinic

Counseling psychology training clinic
Implementation-focused technology development to support MI training and supervision

University of Utah Counseling Center (UCC):
1. Clinical services
2. Clinical training

Study goals:
• Install CORE-MI in UCC IT
• Use to support MI training
• System runs ‘in background’

Implementation framework: User-centered design (UCD)
• Tech goal: Design technology that is maximally implementable
• Clinical goal: Maximize capacity of human therapists to help their clients

UCD processes:
• Needs assessment / task analysis / contextual inquiry
• Design technology for existing workflows and IT
• Iterative design and prototyping / UX testing
We quickly discovered a number of pain points in current system and workflows...

Session recording, naming, and saving error prone

No EHR integration

“Every person named files differently and would often have different video settings making it inconsistent video to video. Videos needed to be saved for legal and flagged for class projects. Videos were deleted by accident. The independent naming process made it more difficult for supervisors to track clients or sessions. Supervisors would have to take notes on file names and times to give structured feedback to the clinician. This made frequent observation cumbersome even for formal supervision.” - Jake Van Epps, PhD, UCC
User-centered design process of designing a cloud platform for counseling training and supervision

Recording platform:
- HIPAA-compliant cloud
- Light EHR integration
- Encrypted video or

Review platform:
- Clinician and supervisor views
- Chat box and session annotation functionality
Mike Smith and John Smith  04/08/17 – 12:45pm

Discussion

MIKE T @ 20:14
Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna.
04/10/17 – 12:45PM

SUPERVISOR
Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna.
04/10/17 – 12:45PM

MIKE T @ 42:54
Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna.
Maecenas sed diam eget risus varius blandit Maecenas sed diam eget risus varius blandit.
04/11/17 – 12:45PM

MIKE T @ 49:54
Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna.
Maecenas sed diam eget risus varius blandit.
04/10/17 – 12:45PM

THERAPIST | Yeah. Yeah I'm here I'm glad to see you yeah I think it's good

CLIENT | See you yeah I think it's good to see you too

THERAPIST | So we can decide together how to do this. Do you want to tell me a little bit about why you're coming in today yeah so.

CLIENT | So we can decide together. Yeah so. There's my twenty first birthday a little bit ago and I drank way too much and was hospitalized and then a couple weeks ago. I was drinking and coming home from the bar and I decide it was OK to drive but I pulled over and I'm going to get a DUI.
Mike S. and John S.  04/08/17 – 12:45pm

OVERALL MI FIDELITY  06/12

MI Adherence  90%
MI Non Adherence  10%
4 Confront
5 Advise
1 Direct

PERCENT TALK TIME
20% Therapist
80% Client

REFLECTION TO QUESTION RATIO
2.8

PERCENT OPEN QUESTIONS: 50%

PERCENT COMPLEX REFLECTIONS: 45%

Comments
SESSION STRENGTHS
Excellent MI Adherence and use of reflections. This is also seen in how much the

AREAS FOR IMPROVEMENT
Let's discuss use of questions, and client's description of fight at minute 6.

ADD REPORT COMMENT

EVOCATION
BASIC:3.5  ADV:4.0
COLLABORATION
BASIC:3.5  ADV:4.0
AUTONOMY/SUPPORT
BASIC:3.5  ADV:4.0
EMPATHY
BASIC:3.5  ADV:4.0
Quality assurance and psychotherapy mechanisms at scale?

Utah Counseling Center

Sessions

<table>
<thead>
<tr>
<th>Sessions</th>
<th>0</th>
<th>250</th>
<th>500</th>
<th>750</th>
<th>1,000</th>
<th>1,250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nov</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Meta-analysis (summary)

Typical research study

Kaiser Permanente...
1.3M counseling sessions in 2017 in southern California

Percentage of Time Therapist is Talking

Ratio of Therapist Reflections to Questions
Current work

Expanding real-world testing and quality metrics:
1. Ongoing research with three implementation partners
2. Patient-centered communication in primary care sessions
3. Cognitive-behavioral therapy pilot work
4. Predicting patient outcomes from linguistic markers (below)

In the language and interaction of the sessions, what is different about these two treatment courses?
Current technology can now scale up the evaluation of evidence-based counseling to support training, supervision, quality assurance, and psychotherapy mechanism research.

Thank you! Questions?

Dave Atkins
dave@lyssn.io

Project updates and project publications available on ResearchGate:
https://www.researchgate.net/project/Machine-learning-based-feedback-for-psychotherapy