Detecting Depression with Audio/Text Sequence Modeling of Interviews

Tuka Alhanai, Mohammad Ghassemi, and James Glass
Massachusetts Institute of Technology
Cambridge MA USA

2nd September, Interspeech 2018

Email: tuka@mit.edu  Website: talhanai.com  Github: github.com/talhanai
How have you been feeling lately?

Healthy: “I’ve been feeling good lately”

Depressed: “*sigh* stressed [um] lately I’ve been really sad and I don’t know why”
Hypothesis: Depression can be modeled using sequences of a spoken interaction.
Hypothesis: Depression can be modeled using sequences of a spoken interaction.
Hypothesis: Depression can be modeled using sequences of a spoken interaction.

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.

Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.

Q: What are some things you don't really like about L.A.?
A: Congestion.
Hypothesis: Depression can be modeled using sequences of a spoken interaction.

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
Hypothesis: Depression can be modeled using sequences of a spoken interaction.

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.

Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.

Q: What are some things you don’t really like about L.A.?
A: Congestion.
Current Solutions

1. Clinical:
   Interview/questionnaire

2. Automated:
   Context-dependent
   Feature engineering
Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated:**
   Context-dependent
   Feature engineering
Current Solutions

1. Clinical: Interview/questionnaire

<table>
<thead>
<tr>
<th>How often during the past 2 weeks were you bothered by...</th>
<th>Not at all</th>
<th>Several days</th>
<th>More than half the days</th>
<th>Nearly every day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Little interest or pleasure in doing things</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2. Feeling down, depressed, or hopeless</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated:**
   Context-dependent
   Feature engineering
Current Solutions

1. Clinical:
   Interview/questionnaire

2. Automated:
   Context-dependent
   Feature engineering
1. **Clinical:** Interview/questionnaire

2. **Automated:** Context-dependent Feature engineering

Current Solutions

1. **Clinical:**
   - Interview/questionnaire

2. **Automated:**
   - Context-dependent
   - Feature engineering

---

Current Solutions

1. **Clinical:** Interview/questionnaire

2. **Automated:** Context-dependent Feature engineering

1. **Clinical:** Interview/questionnaire

2. **Automated:**
   Context-dependent Feature engineering

---

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated:**
   Context-dependent
   Feature engineering

---

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated:**
   Context-dependent
   Feature engineering

---

Current Solutions

1. Clinical:
   Interview/questionnaire

2. Automated:
   Context-dependent
   Feature engineering
Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated:**
   Context-dependent
   Feature engineering

   - keyword: “therapy”
   - “sad”
   - “down”

Current Solutions

1. Clinical:
   Interview/questionnaire

2. Automated:
   Context-dependent
   Feature engineering

Keyword: “therapy”
   “sad”
   “down”


Our Solution

1. Clinical: Interview/questionnaire

2. Automated: Context-dependent Feature engineering
Our Solution

1. Clinical:
   Interview/questionnaire

2. Automated:
   Context-independent
   Feature engineering free
**Q**: What are some things you really like about L.A.?

**A**: I like the weather, I like the opportunities.

**Q**: How easy was it for you to get used to living in L.A.?

**A**: It took a minute, somewhat easy.

**Q**: What are some things you don't really like about L.A.?

**A**: Congestion.

**Q**: What'd you study at school?

**A**: um I took up business and administration.

**Q**: Cool are you still doing that?

**A**: Yeah I am. Here and there, I’m on a break right now but I plan on going back in the uh next semester.

**Q**: What’s your dream job?

**A**: uh probably open up my own business.
Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.

Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.

Q: What are some things you don't really like about L.A.?
A: Congestion.

Q: What'd you study at school?
A: um I took up business and administration.

Q: Cool are you still doing that?
A: Yeah I am. Here and there, I’m on a break right now but I plan on going back in the uh next semester.

Q: What’s your dream job?
A: uh probably open up my own business.
Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
Q: What'd you study at school?
A: um I took up business and administration.
Q: Cool are you still doing that?
A: Yeah I am. Here and there, I’m on a break right now but I plan on going back in the uh next semester.
Q: What’s your dream job?
A: uh probably open up my own business.
Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
Q: What'd you study at school?
A: um I took up business and administration.
Q: Cool are you still doing that?
A: Yeah I am. Here and there, I'm on a break right now but I plan on going back in the uh next semester.
Q: What's your dream job?
A: uh probably open up my own business.
Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
Method

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.

Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.

Q: What are some things you don't really like about L.A.?
A: Congestion.

...
Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

Method

bi-LSTM

1/0 depression
Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.

Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.

Q: What are some things you don't really like about L.A.?
A: Congestion.

...
**Method**

1/0 depression

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.

Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.

Q: What are some things you don't really like about L.A.?
A: Congestion.

...
Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
What do you like about LA?
How have you been feeling lately?
What’s your sleep been like?

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.

Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.

Q: What are some things you don’t really like about L.A.?
A: Congestion.
What do you like about LA?
How have you been feeling lately?
What’s your sleep been like?

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.

Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.

Q: What are some things you don't really like about L.A.?
A: Congestion.
Data

• Corpus: Distress Analysis Interview (DAIC).
Data

- **Corpus:** Distress Analysis Interview (DAIC).
- **Recordings:** 142 (train = 107, test = 35).
Data

• **Corpus:** Distress Analysis Interview (DAIC)
• **Recordings:** 142 (train = 107, test = 35).
• **Structure:** *Wizard-of-Oz* dialogue with 170 unique questions and 8,050 responses.
Data

• **Corpus:** Distress Analysis Interview (DAIC).
• **Recordings:** 142 (train = 107, test = 35).
• **Structure:** *Wizard-of-Oz* dialogue with 170 unique questions and 8,050 responses.
• **Outcome:** binary (28 depressed).
Features

• Audio: \((x \ 279)\)
Features

• **Audio**: (x 279)
  Spectral energy, prosody, and voice quality.
Features

- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

- **Text:** Doc2Vec embeddings of question-response.
  \( \text{dim} = 100 \), \( \text{min-words} = 3 \), \( \text{context-win} = 3 \), \( \text{epochs} = 50 \).
Features

• **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

  ![Audio Spectrum](image)

• **Text:** Doc2Vec embeddings of question-response.
  dim = 100, min-words = 3, context-win = 3, epochs = 50.
Features

• **Audio**: (x 279)
  Spectral energy, prosody, and voice quality.

  ![Audio signal](image)
Features

• **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

  Higher order statistics of response segment.
  \{mean, max, min, med, std, skew, kurt\}
Features

• **Audio**: (x 279)
  Spectral energy, prosody, and voice quality.
  
  Higher order statistics of response segment.
  \{mean, max, min, med, std, skew, kurt\}

• **Text**: (x 100)
Features

• **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.
  
  Higher order statistics of response segment.
  \{mean, max, min, med, std, skew, kurt\}

• **Text:** (x 100)
  Doc2Vec embeddings of question-response.
Features

• **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.
  Higher order statistics of response segment.
  \{mean, max, min, med, std, skew, kurt\}

• **Text:** (x 100)
  Doc2Vec embeddings of question-response.
Features

• **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

Higher order statistics of response segment.
  \{mean, max, min, med, std, skew, kurt\}

• **Text:** (x 100)
  Doc2Vec embeddings of question-response.
  \{dim = 100, min-words = 3, context-win = 3, epochs = 50\}
Results

![Diagram showing F1 Score with baseline (text) comparison.]

Williamson et al.

baseline (text)
Results

- Baseline (text)
- Text
- Audio

F1 Score

Williamson et al.

Context-free
Results

![Bar chart showing F1 scores for baseline, text, and audio models.]

- **Baseline (text)**: Higher F1 score
- **Text**: Moderate F1 score
- **Audio**: Lower F1 score

**Examples**: Williamson et al., Context-free, Weighted
Results

Overfitting on text.

- Williamson et al.
- Context-free
- Weighted
Overfitting on text. Audio generalizes better.
Results

F1 Score

- Williamson et al.
- Context-free
- Weighted
- Sequence

Comparison of F1 scores for different models and modalities.
Results

Better generalization.
Results

Better generalization.
Depression cues exist at relatively longer audio intervals.
Results

The graph shows F1 scores for different models and datasets. The models are labeled as follows:

- **Baseline (text)**
- **Text**
- **Audio**
- **Text + Audio**

The datasets are labeled as:

- **Williamson et al.**
- **Context-free**
- **Weighted**
- **Sequence**

The F1 scores range from 0.4 to 0.8, with each model and dataset having a different color for visual distinction.
Results

Complementary information.
Results

Complementary information.
Conclusion

1. **Sequence Modeling:**
   Improved generalization.
   Better than baseline(s).

2. **Modality Inputs to Model:**
   audio = 30 sequences
   text = 7 sequences
   Depression cues exist at longer speech intervals.

3. **Weighted Model:**
   Overfitting on text.
Future Work

1. Apply technique to larger number of subjects with different conditions (dementia).

2. Infer most predictive segments.

3. Infer patterns being captured (speaking rate? keywords?)
Reference


Github: https://github.com/talhanai/redbud-tree-depression