SPEECH CHARACTERISTICS

Enhancing Speech Dialogue with Emotional Intelligence

Brian Garr, CRO
Brian.Garr@govivace.com
From Mid 90’s to today – How far did we come?

- I can speak my phone number and the credit card number to an IVR (and expect it to get that right)
- Speech transcription tending towards “Human Level” accuracy. Most of the errors are around “uh”/“um” “a”/”the” kind of confusions.
- Dictation and Live closed captioning are real, if somewhat error prone.
Now, speech is part of a dialog

Would you like to order a pizza?

Book me a flight from MIA to ROM

Diagnose the MRI

Copyright 2018 GoVivace Speech Technologies 1/15/2018
What can help us in a dialog?

- Address with the proper salutation
- Offer gender appropriate incentives
- Use in disambiguation algorithms

- Block inappropriate transactions
- Offer age appropriate incentives
- Use in disambiguation algorithms
What can help us in a dialog?

- Use colloquial jargon
  - Pop vs Soda
  - Ridy-bob vs seesaw
  - Scaper vs critter

- Transfer call to supervisor
- Offer incentives
- Use in disambiguation algorithms
Smarter systems for today

SPEAKER VERIFICATION
SPEAKER IDENTIFICATION

CHARACTERISTICS
Age
Accent

BETTER OUTCOMES?

Copyright 2018 GoVivace Speech Technologies
Emotion Detection

- Paul Elkman – basic emotions
  - Happy
  - Sad
  - Angry
  - Disgust
  - Fear
  - Surprise

- These emotions are universal across cultures. Facial expressions change similarly as these emotions occur.
Convolutional Neural Networks

- Sparse Connectivity – local receptive field
- Shared Weights – same filter applied across the signal
- Pooling for dimension reduction
Let’s analyze some examples

- We will play an audio example
- Audience will need to guess the following
  - Gender
  - Language - English, Hindi, Thai, German, French, Russian, Arabic, …
  - Age
  - Accent (only for English) - Eastern, NY, CA, TX, Canada, Germany, China, India, France, Russian
  - Level of Arousal (or Neutrality)
- Compete with computer estimates
The guessing game
### The guessing game

<table>
<thead>
<tr>
<th>M/F</th>
<th>Lang</th>
<th>Acc 2</th>
<th>Acc 3</th>
<th>Acc 4</th>
<th>Age</th>
<th>EState</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real M</td>
<td>En</td>
<td>TX</td>
<td></td>
<td></td>
<td>25</td>
<td>NA</td>
</tr>
<tr>
<td>Est M</td>
<td>En</td>
<td>TX .78</td>
<td>IND .6</td>
<td>East 0.59</td>
<td>CDN 0.49</td>
<td>30</td>
</tr>
</tbody>
</table>
The guessing game

<table>
<thead>
<tr>
<th></th>
<th>M/F</th>
<th>Language</th>
<th>Acc 2</th>
<th>Acc 3</th>
<th>Acc 4</th>
<th>Age</th>
<th>EState</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>M</td>
<td>En</td>
<td>TX</td>
<td></td>
<td></td>
<td>25</td>
<td>NA</td>
</tr>
<tr>
<td>Est</td>
<td>M</td>
<td>En</td>
<td>TX</td>
<td>IND</td>
<td>East</td>
<td>30</td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.78</td>
<td>0.59</td>
<td></td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>M/F</td>
<td>Language</td>
<td>Acc 2</td>
<td>Acc 3</td>
<td>Acc 4</td>
<td>Age</td>
<td>EState</td>
</tr>
<tr>
<td>Real</td>
<td>M</td>
<td>En</td>
<td>can</td>
<td></td>
<td></td>
<td>47</td>
<td>NA</td>
</tr>
<tr>
<td>Est</td>
<td>M</td>
<td>En</td>
<td>can</td>
<td>chine</td>
<td>CA/</td>
<td>46</td>
<td>Dominant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.68</td>
<td>a</td>
<td>0.56</td>
<td></td>
<td>0.739</td>
</tr>
</tbody>
</table>
## The guessing game

<table>
<thead>
<tr>
<th>M/F</th>
<th>Lang</th>
<th>Acc 2</th>
<th>Acc 3</th>
<th>Acc 4</th>
<th>Age</th>
<th>EState</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>M</td>
<td>En</td>
<td>TX</td>
<td></td>
<td>25</td>
<td>NA</td>
</tr>
<tr>
<td>Est</td>
<td>M</td>
<td>En</td>
<td>TX</td>
<td>IND</td>
<td>East</td>
<td>CDN</td>
</tr>
<tr>
<td>Real</td>
<td>M</td>
<td>En</td>
<td>can</td>
<td></td>
<td>47</td>
<td>NA</td>
</tr>
<tr>
<td>Est</td>
<td>M</td>
<td>En</td>
<td>can</td>
<td>china</td>
<td>CA/ TX/</td>
<td>46</td>
</tr>
<tr>
<td>Real</td>
<td>F</td>
<td>En</td>
<td>Chin</td>
<td></td>
<td>30</td>
<td>NA</td>
</tr>
<tr>
<td>Est</td>
<td>F</td>
<td>En</td>
<td>Chin</td>
<td>RU/</td>
<td>TX/ NY/</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.71</td>
<td>0.58</td>
<td>0.51</td>
</tr>
</tbody>
</table>
The Complete Conversation

Customer Dialog

- Speech Recognition
- Gender Recognition
- Speaker Identification
- Age Identification
- Geo Location
- Emotional State Recognition
Questions?