A good presenter creates an emotional roller-coaster – better zoom presentations through face emotion recognition

Jiachen Sun\textsuperscript{1,2}, Yucong Lin\textsuperscript{3}
1. MIT Center for Collective Intelligence, MA, USA
2. Sun Yat-sen University, Guangzhou, China
3. Center for Statistical Science, Tsinghua University, Beijing, China

* Cooperated with Peter Gloor and Jannik Roessler
How to make a good public presentation

- The content of the speech (Nikitina et. al., 2011; Chollet et. al., 2015)
- Interaction with the audience (Braga D & Marques M A, 2004)
- Sound quality (Wichmann A, 2002)
- Fluency (Scherer, S et. al., 2012)
- nonverbal communication (Strangert and Gustafson, 2018; S Scherer et. al. 2012)
- **motivating listeners' emotion**
  - Definition: The ability of motivating listeners' emotion refers to the ability of the speaker affect the emotions of the audience through verbal communication skills and non-verbal communication skills at different stages of the speech, thereby enhancing the effect of the speech.
How to analysis the effect of the skill of motivating listeners' emotion on public speaking

- The direct idea is the randomized controlled test

  - experiment group
  - control group

  - Same context
  - Same speaker
  - Similar characteristic in listener group
  - Different ability of motivating listeners' emotion

  Y: Speaking quality
  X: The ability of motivating listeners' emotion

- However, there are some difficulty to apply the experiment:

  1. How to quantify the ability of motivating listeners’ emotion? (Related to sounds, accents, fluency …)

  2. How to control the features affecting to Y (speaking quality)?
    1. E.g. An extremely bad context will cause low speaking quality whatever the ability of motivating listeners’ emotion
Data Pre-processing

Step 1  Zoom Record

Step 2  Video to Frame (per second)

Step 3  Frame segmentation

Step 4  Emotion Recognition

Neutral
Happy
Sad
Surprise
Angry
Fear

Deep Learning Model
Data

Collaborative Innovation Networks

- **Participants**
  - 35 students from University of Cologne and University of Bamberg.
  - A total of 8 teams

- **Virtual Meeting (Zoom)**
  - held every two weeks for a total of 6 times
  - 15-minute Presentation for each team

- **Presentation Rating**
  - Poll for Audience
  - Take the mean value from all audiences as the collective score Y for a presentation
Building AI-classifier

- **6 Datasets:** FER2013 + AffectNet + CK + BU + Google + JAFFE
- **6 emotions:** Neutral, Happy, Sad, Surprise, Fear, Anger
- **Different deep neural networks**
  
<table>
<thead>
<tr>
<th></th>
<th>VGG13</th>
<th>VGG16</th>
<th>VGG19</th>
<th>Xception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.826</td>
<td><strong>0.840</strong></td>
<td>0.836</td>
<td>0.837</td>
</tr>
</tbody>
</table>

- **Best model:** VGG16
  - Used to predict the emotion of Zoom data
Temporal Sequence of 6 Emotions at a Presentation

Time (second)

Neutral
Happy
Surprise
Sad
Angry
Fear

Number of emotions
**Speaker’s emotion and Audience’s emotion**

**Correlation Matrix (N=41)**

1. The happier the speaker is, the happier and less neutral the audience is.
2. The more neutral speaker is, the less surprise the audience is.
## Correlation between Emotion and Presentation score

<table>
<thead>
<tr>
<th>Significant variables</th>
<th>Pearson-r</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dev. of happy</td>
<td>0.73</td>
<td>7e-08</td>
</tr>
<tr>
<td>Dev. of neutral</td>
<td>0.50</td>
<td>8e-04</td>
</tr>
<tr>
<td>Dev. of fear</td>
<td>0.34</td>
<td>0.025</td>
</tr>
<tr>
<td>Sum of happy</td>
<td>0.55</td>
<td>2e-04</td>
</tr>
<tr>
<td>Sum of neutral</td>
<td>-0.44</td>
<td>0.004</td>
</tr>
<tr>
<td>Sum of fear</td>
<td>0.38</td>
<td>0.01</td>
</tr>
<tr>
<td>Density of happy</td>
<td>0.44</td>
<td>0.004</td>
</tr>
<tr>
<td>Speaker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of happy</td>
<td>0.35</td>
<td>0.026</td>
</tr>
</tbody>
</table>
Pearson r = 0.73
P = 7.0 \times 10^{-8}
Predicting Presentation’s Score
(Ordinary Least Squares regression)

(a) Predictor: dev_happy, dev_fear

(b) Pearson r = 0.75
P = 1.3 × 10⁻⁸
Conclusion

• An Empirical Study of Zoom Presentation through Facial Emotion Recognition

• Identification of useful features from Emotion of Speaker and Audience

• Prediction of presentation performance using emotional feature
Reference