RideSafe -- Detecting Sexual Harassment in Rideshares

Shikhar Sahuja, Robin Cohen
{s2sahuja, rcohen}@uwaterloo.ca
David R. Cheriton School of Computer Science
University of Waterloo
Rideshare Market

- 65 countries and 600 cities
- 14 million trips each day
More people use Uber than the total population of these three countries:

- New Zealand (4,783,063)
- Denmark (5,771,876)
- Singapore (5,804,337)
2002:
“Don’t get in the car with strangers”
“Don’t meet people from the internet”

2018:
Uber
Uber reports more than 3,000 sexual assaults during U.S. rides in 2018

Uber will pay $4.4 million to settle a federal probe into sexual harassment and retaliation

The charges stem from an investigation that started in 2017

By Andrew J. Hawkins | @andyjayhawk | Dec 19, 2019, 11:56am EST
What can we do about this?

- Applications of AI; multimedia processing.
- Voice recordings from rideshares analyzed.
- Detecting fear in female passengers voice.
- Cases directly forwarded to authorities and rideshare company.
- Relevant Work: Speech Processing, Feature Extraction, Model Selection
Technical Solution
Dataset

- Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)
- The RAVDESS contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent.
- Speech emotions includes calm, happy, sad, angry, fearful, surprise, and disgust expressions.
Data Preprocessing

- Moving audio to frequency domain using Mel-Frequency Cepstral Coefficients (MFCCs)
- MFCC that try to map phonemes as the shape of the vocal tract that generates the sound.
- More precisely, maps audio to short term power spectrum that allows computers to understand sound the way we do.
- Pitch considered in the models.
Support Vector Machines

- Detects fear in voice by treating fear as positive class and others as negative.
- Applying to voice recordings (time-based), MFCC average.
- F1 score of 91%.
- Treat MFCC as an image for CNN.
- Convolve over parts of the MFCC “image” to train the network.
- F1 score of 95%. 

**CNNs**
LSTM

- Stacked LSTM model
- Applying to voice recordings (time-based), MFCC average.
- F1 score of 90%.
Results and Discussion

- Minimize FNR to avoid cases from slipping through. Could be life and death.
- High FPR can be addressed with additional checks.
- Hybrid SVM and 1D-CNN model minimizes both.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>FPR</th>
<th>FNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>100%</td>
<td>84%</td>
<td>91%</td>
<td>0%</td>
<td>16%</td>
</tr>
<tr>
<td>LSTM</td>
<td>87%</td>
<td>92%</td>
<td>90%</td>
<td>43%</td>
<td>7%</td>
</tr>
<tr>
<td>1D-CNN</td>
<td>92%</td>
<td>98%</td>
<td>95%</td>
<td>25%</td>
<td>1%</td>
</tr>
</tbody>
</table>
- Fear can be attributed to other reasons.
  - More false positives.
- Response to trauma might not always be emotional.
  - Victims might not react to avoid confrontation.
- Recording audio, even in rideshares, can imply privacy violations for many.
  - Explicit permission.
- Creates a system of distrust with the drivers.
  - Need of the hour, unfortunately.
- Could include sensitive information about the rider and/or driver.
  - Access protocols regarding how the audio is stored and who gets access to it.
  - Raw audio only accessible to very few.
  - Derived features such as MFCC, Spectogram, Pitch, Frequency information saved in databases.
Future Work and Conclusion

- Incorporate data for other at-risk groups such as transgendered.
- Integrate text analysis for the voice recordings to further identify harassment.
- Valuable current application where AI can assist.
- The application can help deter cases of sexual harassment and lead to a safer rideshare experience.
Algorithm to compute MFCC

1. Frame the signal into short frames.
2. For each frame calculate the periodogram estimate (Discrete Fourier Transform + periodogram-based power spectral estimate) of the power spectrum.
3. Apply the mel filterbank to the power spectra, sum the energy in each filter.
4. Take the logarithm of all filterbank energies.
5. Take the Discrete Cosine Transform of the log filterbank energies.
6. Keep Discrete Cosine Transform coefficients 2-13, discard the rest.
Questions?