Speech Emotion Recognition with a Reject Option

Kusha Sridhar, Carlos Busso

Erik Jonsson School of Engineering & Computer Science at the University of Texas at Dallas, Richardson, Texas 75080, USA

Motivation

- Speech emotion recognition is a hard problem
- Prediction are not always reliable
- Abstaining from prediction when in doubt can increase the reliability of a system
- Selective classification on images have led to very low error rate (2%) for a test coverage of 60%
- The key challenge is to define mechanisms to quantify reliability to accept or reject an instance
  - e.g., Apply threshold on softmax output

Our Work

- SER system with a reject option
  - Accept or reject a sample based on the confidence of the classifier
  - Defined thresholds to interpret the confidence

- Goal is to improve the classifier performance while maintaining a high test coverage

Database and Features

The MSP-Podcast Corpus

- Emotionally rich speaking turns from speakers appearing in various podcasts (2.75s – 11s)
- Annotated for primary and secondary emotions (crowdsourcing)
- V1.4: 33,262 utterances with emotional labels (56h 29m)
  - Train set: 19,707 segments
  - Test set: 9,255 segments from 50 speakers
  - Validation set: 4,300 segments from 30 speakers
- Five-class problem
  - Happiness, Neutral, Sadness, Anger, Disgust
- Eight-class problem
  - Happiness, Neutral, Sadness, Anger, Disgust, Surprised, Contempt, Fear

Acoustic Features

- Interspeech 2013 Computational Paralinguistic Challenge
  - feature set (6,373 features extracted with OpenSmile)

reject option for speech emotion recognition

Reject Option for SER

Criterion 1:

- Threshold on the neuronal activations
- Selective guaranteed risk (SGR) algorithm
- Learn optimal risk bound on the classifier
- Threshold on softmax outputs to achieve a desired error rate with high confidence

\[ r(f, g|S_m) = \frac{1}{m} \sum_{i=1}^{m} I(f(x_i), y_i)g(x_i) \]

\[ Pr(S_m \{ r(f, g|S_m) < r^* \} > 99.99\% \]

\[ \phi(f, g|S_m) \leq \frac{1}{m} \sum_{i=1}^{m} g(x_i) \]

Criterion 2:

- Threshold on the difference between the two highest prediction values
- Large difference \(\rightarrow\) clear prediction \(\rightarrow\) accept

Optimization

- Empirical risk of classifier using SGR algorithm
- F1-Score

Architecture

- Two layers
- 1,024 nodes
- ReLU activation
- ADAM optimizer

Task

- Categorical emotion recognition

Analysis & Conclusion

Inter-Evaluator agreement of accepted/rejected samples

<table>
<thead>
<tr>
<th>Inter-evaluator agreement (Fleiss Kappa)</th>
<th>Coverage (%)</th>
<th>Accepted samples</th>
<th>Rejected samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard labels (5-class)</td>
<td>100</td>
<td>0.2642</td>
<td>0.2590</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0.2723</td>
<td>0.2651</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.2897</td>
<td>0.2668</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.3080</td>
<td>0.2633</td>
</tr>
<tr>
<td>Soft labels (8-class)</td>
<td>100</td>
<td>0.2680</td>
<td>0.2450</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>0.2723</td>
<td>0.2460</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.2842</td>
<td>0.2496</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.2983</td>
<td>0.2563</td>
</tr>
</tbody>
</table>

Observations

- Lower inter-evaluator agreement for rejected samples

Conclusions

- The reject option is a valuable feature, increasing the confidence in a SER system
- Improvement in performance without compromising much on the coverage in the test set

References:


This work was supported by NSF under Grant CNS-1823166 and CAREER Grant IIS-1453781