A reliability-aware model for intelligibility classification in pathological speech

Naveen Kumar, Shrikanth Narayanan

Signal Analysis and Interpretation Lab
University of Southern California
1. Pathological Speech
2. Defining reliability
3. Reliability-aware classification model
4. Results
5. Summary
Pathological Speech

- Atypicality resulting from disease or surgery of the vocal tract
- Reduced speech intelligibility
- Decrease of intelligibility might be perceived by different factors
NKI CCRT Speech Corpus

- Speech Intelligibility before and after treatment.
- Released during Interspeech 2012 speaker trait challenge\[^1\]
- 2385 sentence level utterances
- Ratings thresholded to
  - intelligible (I)
  - non-intelligible (NI)

Feature Subsets

Utterance level feature subsystems adapted from [2]

- **Prosody** (6)
  - Pitch - L0 norm, polynomial fit, variance
- **Pronunciation** (2)
  - CMN 39 dim. MFCC, phone duration
- **Voice Quality** (5)
  - HNR, Jitter, Shimmer

---

How was the label $Y$ assigned given features $X$? (discriminative)

Defining Reliability

reliable ($R=1$)  
unreliable ($R=0$)
How was the label $Y$ assigned given features $X$? \textit{(discriminative)}

Why model label reliability?

- human annotations are inherently subjective
- noisy features on some samples

Reliability of labels

- reliable ($R=1$)
- unreliable ($R=0$)
Reliability formulation

\[ Pr(X, Y | R) = Pr(X, Y ; \Theta)^R [Pr(X ; \Phi) Pr(Y ; \Phi)]^{1-R} \]

- \( R=\{0,1\} \): reliable at random model
- \( \Theta \): data dependent reliable model
- \( \Phi \): data independent unreliable model

Latent reliability \( R \) controls dependence between data \( X \) and label \( Y \).
Discriminative reliability assumption

• **unreliable**: Label $Y$ generated independent of data
• **reliable**: label $Y$ generated according to a data-dependent model

\[
Pr(Y|X, R) = \underbrace{Pr(Y|X; \Theta)^R}_{\text{reliable}} \quad \underbrace{Pr(Y; \Phi)^{1-R}}_{\text{unreliable}}
\]
Discriminative reliability assumption

- **unreliable**: Label $Y$ generated independent of data
- **reliable**: label $Y$ generated according to a data-dependent model

\[
Pr(Y|X, R) = \underbrace{Pr(Y|X; \Theta)^R}_{\text{reliable}} \cdot \underbrace{Pr(Y; \Phi)^{1-R}}_{\text{unreliable}}
\]

- **data-dependent reliability**

\[
Pr(Y, R|X) = Pr(R|X) \cdot Pr(Y|X; \Theta)^R \cdot Pr(Y; \Phi)^{1-R}
\]

- **mixture of experts model**
| Reliable | Maximum-Entropy (softmax) | \( Pr(Y_i = k| X_i; \Theta) = \frac{e^{W_k^T X_i}}{\sum_{j=1}^{K} e^{W_j^T X_i}} \) | \( \Psi_{ik}(W) \) |
|---|---|---|---|
| Unreliable | Multinomial | \( Pr(Y_i = k; \Phi) = \lambda_k \) | \( \lambda_k \) |
| Reliability Model | Logistic Regression | \( Pr(R_i = 1| X_i) = \sigma(r^T X) \) | \( \rho_i \) |
ML estimation via EM algorithm

- **E-step**
  \[ \gamma_i \]
  \[
  Pr(R_i = 1|X_i, Y_i) = \frac{Pr(Y_i|R_i = 1, X_i) \ Pr(R_i = 1|X_i)}{Pr(Y_i|X_i)}
  \]

- **M step**
  estimate parameters \( W^* \, \lambda^* \, r^* \) by maximizing expected data log-likelihood.
Data log-likelihood

\[ r^* \]

\[ W^* \]

\[ r^* \]

\[ N \]

\[ i=1 \]

\[ \gamma_i \ln \rho_i + (1 - \gamma_i) \ln(1 - \rho_i) \]

\[ r^* \]

\[ N \]

\[ K \]

\[ i=1 \]

\[ k=1 \]

\[ Y_{ik}[\gamma_i \ln \Psi_{ik} + (1 - \gamma_i)\ln \lambda_k] \]

data dependent reliability model

reliable and unreliable models
Data log-likelihood

\[\sum_{i=1}^{N} \gamma_i \ln \rho_i + (1 - \gamma_i) \ln(1 - \rho_i)\]

weighted maxent with weights \(\gamma_i\)
### Parameter updates and Inference

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Formula</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_k )</td>
<td>( \frac{\sum_{i=1}^{N} Y_{ik}(1 - \gamma_i)}{\sum_{i=1}^{N} Y_{ik}} )</td>
<td>one step</td>
</tr>
<tr>
<td>( r )</td>
<td>logistic</td>
<td>gradient ascent (L-BFGS)</td>
</tr>
<tr>
<td>( W )</td>
<td>weighted-maxent</td>
<td>gradient ascent (L-BFGS)</td>
</tr>
</tbody>
</table>
Parameter updates and Inference

| $\lambda_k$ | $\frac{\sum_{i=1}^{N} Y_{ik}(1 - \gamma_i)}{\sum_{i=1}^{N} Y_{ik}}$ | one step |
| r | logistic | gradient ascent (L-BFGS) |
| $W$ | weighted-maxent | gradient ascent (L-BFGS) |

Inference of class labels

\[ Pr(Y = k|Z) = \Psi_k^{\star}(W^*, Z)\sigma(r^T Z) + \lambda_k^\star[1 - \sigma(r^T Z)] \]

test feature
Experiment: Intelligibility classification

- 5 fold cross validation
- Baseline: Reliability blind classifier, always assumes R=1

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Logistic regression</th>
<th>Reliability aware</th>
</tr>
</thead>
<tbody>
<tr>
<td>voice quality</td>
<td>58.2</td>
<td>59.8</td>
</tr>
<tr>
<td>prosody</td>
<td>67.1</td>
<td>66.7</td>
</tr>
<tr>
<td>pronunciation</td>
<td>55.1</td>
<td>56.2</td>
</tr>
<tr>
<td>feature fusion</td>
<td>68.0</td>
<td>67.8</td>
</tr>
</tbody>
</table>
• $R \sim \text{Bernoulli}(p)$

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Logistic regression</th>
<th>Reliability aware</th>
<th>Avg. Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>voice quality</td>
<td>58.2</td>
<td>59.8</td>
<td>0.43</td>
</tr>
<tr>
<td>prosody</td>
<td>67.1</td>
<td>66.7</td>
<td>0.73</td>
</tr>
<tr>
<td>pronunciation</td>
<td>55.1</td>
<td>56.2</td>
<td>0.16</td>
</tr>
<tr>
<td>feature fusion</td>
<td>68.0</td>
<td>67.8</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Reliability aware model improves classification when feature set is less reliable
Summary

Pros
• discriminative modeling allows for reliable parameter estimation
• learns regions in feature space where annotations are more reliable

Cons
• linear class boundary for reliability in feature space may not be ideal
• model is unable to combine reliable information from different feature subsets
Questions?