HumanACGAN: conditional generative adversarial network with human-based auxiliary classifier and its evaluation in phoneme perception

1. SYNOPSIS

1. Proposes a generative adversarial network (GAN) that represents conditional human acceptable distributions.
2. Demonstrates that the proposed algorithm successfully trains a generator.

2. RESEARCH BACKGROUND

<table>
<thead>
<tr>
<th>Human-acceptable distribution</th>
<th>Modeling method</th>
<th>Conditional GAN</th>
<th>Unconditional ACGAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humans can accept outside of real data</td>
<td>Real-data distribution</td>
<td>HumanACGAN (ours)</td>
<td>HumanACGAN</td>
</tr>
</tbody>
</table>

3. RELATED WORKS

<table>
<thead>
<tr>
<th>Generator $G$</th>
<th>Discriminator $D_S$</th>
<th>Classifier $D_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAN [1]</td>
<td>DNN</td>
<td>DNN</td>
</tr>
<tr>
<td>HumanGAN [2]</td>
<td>DNN</td>
<td>Humans</td>
</tr>
<tr>
<td>ACGAN [3]</td>
<td>DNN</td>
<td>DNN</td>
</tr>
</tbody>
</table>

**GAN [1]**

Minimizing $L_S = \sum \log(1 - D_S(\hat{x}_n))$ during generator training. Gradient is calculated analytically.

**HumanGAN [2]**

Maximizing $L_S = \sum D_S(\hat{x}_n)$ during generator training. Estimating gradient using Natural evolution strategy (NES) [4].

**ACGAN [3]**

Maximizing $-L_S + \lambda L_C$ during generator training. $L_C = \sum \log(D_C(x_n, c_n)) + \sum \log(D_C(\hat{x}_n, c_n))$. Gradients are calculated analytically.

4. PROPOSED METHOD: HumanACGAN

Maximizing $L_S + \lambda L_C$ during generator training.

$$L_S + \lambda L_C = \sum D_S(\hat{x}_n, c_n)$$

Estimating gradient using NES.

$$\frac{\partial L_S + \lambda L_C}{\partial \hat{x}_n} = \frac{1}{2\sigma} \sum_{i=1}^{R} \Delta D_S(\hat{x}_n^{(i)}; c_n) \cdot \Delta x_n^{(i)}$$

Backpropagation using $\Delta x^{(r)}$ and $\Delta x^{(r')}$

5. EXPERIMENTAL EVALUATIONS

**Experimental conditions**

- **Corpus**: JVPD [5]
- **Analysis/synthesis**: WORLD [6]
- **Optimization**: Stochastic gradient descent
- **Data space**: Applying PCA (principal component analysis) to log spectral envelope and using 1st and 2nd PC as data space

**Experiment 1: estimation of gradient**

- **Method**: Preparing perturbed data $\hat{x} + \Delta x$ and $\hat{x} - \Delta x$
- **Humans’ tasks**: Synthesizing speech waveform from $\hat{x} + \Delta x$ and $\hat{x} - \Delta x$
- **Results**: Gradients pointed to darker (i.e., higher posterior) zones.

**Experiment 2: increase of objective function**

- **Method**: Training generator for 4 iterations using estimated gradient
- **Results**: Prior probabilities of naturalness and class acceptability increased.

References: