Fixed that For You: Generating Contrastive Claims with Semantic Edits

Chris Hidey and Kathleen McKeown
Columbia University

Motivation and Goal

Contrast is important in argumentation for understanding differences in opinions.

Explicit negation limits diversity of responses (Bilu et al., 2015)

Instead generate contrastive responses with semantically relevant replacements

Data Collection

Reddit – mined comment pairs with FTFY (fixed that for you)

Get employers out of the business, pass universal single-payer healthcare. => Get employers out of the business, deregulate and allow cross-state competition. FTFY

AMT – worker-generated response given Reddit claim

1) Data from 10 most popular categories

   e.g. Movies, Sports, Video Games

2) Qualified annotators that passed a knowledge quiz

   Steven Spielberg is the greatest director of all time.
   a) Paul McCartney
   b) Stanley Kubrick

Annotated Categories for Reddit

• Contrast (51.4 %)

   Nintendo is the only company that puts customers over profits => Rockstar is the only company

   Ted Cruz for president => Zodiac killer for president

• Jokes (27 %)

   This python library really piques my interest => *py*quest

• Corrections (21.2 %)

   This peaks my interest => piques

Data Classification

Model Features

• Character-based (Jaccard similarity, edit distance, etc.)

• GloVe embeddings (mean of span, cosine similarity, etc.)

• Vocabulary (known word, non-ASCII, etc.)

Model for Contrastive Claim Generation

Copy Prediction

Constrained Decoding

Bernie for president

Hillary for president

Counter Features

Topic Features

Automatic Evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>Novel</th>
<th>Reddit BLEU-4</th>
<th>Reddit % Match</th>
<th>AMT BLEU-4</th>
<th>AMT % Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2Seq w/ Attn.</td>
<td>3.3</td>
<td>7.3</td>
<td>0.7</td>
<td>25.8</td>
<td>0.7</td>
</tr>
<tr>
<td>CTR</td>
<td>16.2</td>
<td>8.5</td>
<td>2.0</td>
<td>27.7</td>
<td>2.4</td>
</tr>
<tr>
<td>CTR+TOPIC+CPY</td>
<td>16.3</td>
<td>9.62</td>
<td>2.0</td>
<td>31.2</td>
<td>3.8</td>
</tr>
<tr>
<td>Seq2Seq w/ Attn.</td>
<td>79.9</td>
<td>18.8</td>
<td>4.6</td>
<td>40.1</td>
<td>10.1</td>
</tr>
<tr>
<td>CTR</td>
<td>89.7</td>
<td>22.6</td>
<td>4.7</td>
<td>47.6</td>
<td>12.6</td>
</tr>
<tr>
<td>CTR+TOPIC+CPY</td>
<td>90.5</td>
<td>23.1</td>
<td>4.8</td>
<td>50.1</td>
<td>14.9</td>
</tr>
</tbody>
</table>

Human Evaluation

<table>
<thead>
<tr>
<th>Scale: 1 to 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Unconstrained Baseline</td>
</tr>
<tr>
<td>Constrained Baseline</td>
</tr>
<tr>
<td>Constrained CTR+TOPIC+CPY</td>
</tr>
</tbody>
</table>

Model Input and Output

Our Contributions

• Distant-labeled and annotated contrastive claim pairs

• Methods to identify and filter noisy data

• Neural models to generate contrastive claims

Blog post: [Fixed that For You: Generating Contrastive Claims with Semantic Edits](https://example.com/blog-post)