Summary

Task: We address the low-diversity problem of Sequence-to-Sequence (Seq2Seq) based chatbots – the problem of frequently generating dull responses like “I don’t know” or “I’m sorry”.

Method: Other than previous diagnoses, we believe this problem is also caused by frequent tokens. Therefore, we propose a new loss function to counter the effect of different token frequencies.

Main contributions:
- We show that frequency variance of different tokens can cause model over-confidence and low response diversity.
- We propose a Frequency-Aware Cross-Entropy (FACE) loss function to balance per-token training loss, which alleviates model over-confidence and, hence, improves response diversity.
- We investigate two token frequency calculation methods and corresponding frequency-based weighting mechanisms for FACE.

Problem Analysis

Model over-confidence and low-diversity are statistical and empirical symptoms of the same problem: imbalance of training losses caused by token frequency variances. What is model over-confidence?

Frequent tokens in training data will result in frequent tokens in model outputs. Why does it influence the response diversity?

Frequent Tokens + Language Model = Frequent Responses

Frequency-Aware Cross-Entropy Loss

How do we make token generation more balanced? We balance model training loss: that’s how model is directly influenced. FACE loss function:

\[ \text{FACE}(t) = \exp(-\text{CE}(t) \cdot c) \]

Experimental Results

Automatic & human evaluations on the Twitter dataset:

Some examples:

<table>
<thead>
<tr>
<th>Model</th>
<th>Human evaluation results (p-value: &lt; 0.05)</th>
<th>Human evaluation results (p-value: &lt; 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACE vs CMHAM</td>
<td>FACE vs MHAM</td>
<td>FACE vs MHAM</td>
</tr>
<tr>
<td>FACE vs Mii-bidi</td>
<td>FACE vs Mii-bidi</td>
<td>FACE vs Mii-bidi</td>
</tr>
<tr>
<td>FACE vs Mii-antiLM</td>
<td>FACE vs Mii-bidi</td>
<td>FACE vs Mii-bidi</td>
</tr>
<tr>
<td>FACE vs Seq2Seq</td>
<td>FACE vs Seq2Seq</td>
<td>FACE vs Seq2Seq</td>
</tr>
</tbody>
</table>

Conclusion

- FACE loss function can effectively improve diversity and quality of responses.
- FACE achieves improvements with minimum modifications to original Seq2Seq model, which makes it flexible to extend.
- A limitation of FACE is that learning procedure is not as stable as cross-entropy, which increases difficulty of training.

Source Code & Twitter