Tell Me How to Ask Again: Question Data Augmentation with Controllable Rewriting in Continuous Space

Dayiheng Liu, Yeyun Gong, Jie Fu, Yu Yan, Jiusheng Chen, Jiancheng Lv, Nan Duan, Ming Zhou

EMNLP 2020
Task

- **Question answering:**
  - Given a question
  - Given a document
  - Find the span in the document in which answer is provided
  - Binary classification if the question is answerable or not

- **They look for a method to augment the questions:**
  - To have the same answer span
  - To be answerable/unanswerable from the document
  - To be similar to the original question
Pre-trained Language Model based MRC Model

- A BERT-based model to classify the input question is answerable or not and to find the answer span

\[
E^q, E^d = \text{BertEmbedding}(q, d),
\]

\[
P_a(\text{is-answerable}) = \text{Sigmoid}(C W^T_c + b_c),
\]

\[
P_s(i = \langle \text{start} \rangle) = \text{Sigmoid}(T^d_i W^T_s + b_s),
\]

\[
P_e(i = \langle \text{end} \rangle) = \text{Sigmoid}(T^d_i W^T_e + b_e),
\]

\[
\mathcal{L}_{\text{mrc}} = \lambda \mathcal{L}_a(t) + \mathcal{L}_s(s) + \mathcal{L}_e(e),
\]

\[
= -\lambda \log P_a(t) - \log P_s(s) - \log P_e(e),
\]
Transformer-based Autoencoder

- Encode the question
  - Use the same embedding as the pre-trained model
- Compute a vector representation of the input question
- Decode the question vector using a Decoder

\[ H_{enc} = \text{TransformerEncoder}(q), \]
\[ z = \text{Sum}(\text{GRU}(H_{enc})), \]
\[ \hat{q} = \text{TransformerDecoder}(z). \]
Rewriting Question with Gradient-based Optimization

- Three objectives for rewriting:
  - Be unanswerable or have the same span
  - Should not be trapped by local optimum
  - Should be similar to Q

- Unanswerable question: \( E_{q'} = E_q - \eta(\nabla_{E_q} L_a(t')) \)
- Same Span: \( E_{q'} = E_q - \eta(\nabla_{E_q}(\lambda L_a(t) + L_s(s) + L_e(e))) \)
- Update step-size for avoiding local optimum
- Use unigram overlap rate for choosing similar questions: \( J(q, q') = \frac{\text{count}(w_q \cap w_{q'})}{\text{count}(w_q \cup w_{q'})} \)
Algorithm 1 Question Rewriting with Gradient-based Optimization.

**Input:** Data tuple \((q, d, s, e, t)\); Original question embedding \(E^q\); pre-trained MRC model and Transformer autoencoder; A set of step size \(S_\eta = \{\eta_i\}\); Step size decay coefficient \(\beta_s\); the target answerable or unanswerable label \(t'\); Threshold \(\beta_t, \beta_a, \beta_b\);

**Output:** a set of new answerable and unanswerable question data tuples \(\mathcal{D}' = \{(q', d, s, e, t'), ..., (q', d, s, e, t)\}\);

1: \(\mathcal{D}' = \{\}\);
2: for each \(\eta \in S_\eta\) do
3:     for max-steps do
4:         revise \(E^{q'}\) by Eq. (10) or Eq. (9)
5:         \(q' = \text{TransformerAutoencoder} (E^{q'})\)
6:         if \(P_a(t') > \beta_t\) and \(\mathcal{J}(q, q') \in [\beta_a, \beta_b]\) then
7:             add \((q', d, s, e, t')\) to \(\mathcal{D}'\);
8:         end if
9:         \(\eta = \beta_s \eta\);
10:     end for
11: end for
12: return \(\mathcal{D}'\);
## Results

<table>
<thead>
<tr>
<th>Methods</th>
<th>EM</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{BERT}_{\text{large}}$ (Devlin et al., 2018) (original)</td>
<td>78.7</td>
<td>81.9</td>
</tr>
<tr>
<td>+ EDA (Wei and Zou, 2019)</td>
<td>78.3</td>
<td>81.6</td>
</tr>
<tr>
<td>+ Back-Translation (Yu et al., 2018)</td>
<td>77.9</td>
<td>81.2</td>
</tr>
<tr>
<td>+ Text-VAE (Liu et al., 2019a)</td>
<td>75.3</td>
<td>78.6</td>
</tr>
<tr>
<td>+ AE with Noise</td>
<td>76.7</td>
<td>79.8</td>
</tr>
<tr>
<td>+ 3M synth (Alberti et al., 2019)</td>
<td>80.1</td>
<td>82.8</td>
</tr>
<tr>
<td>+ UNANSQ (Zhu et al., 2019)</td>
<td>80.0</td>
<td>83.0</td>
</tr>
<tr>
<td>+ CRQDA (ours)</td>
<td><strong>80.6</strong></td>
<td><strong>83.3</strong></td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Methods</th>
<th>EM</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT\textsubscript{base}</td>
<td>73.7</td>
<td>76.3</td>
</tr>
<tr>
<td>+ CRQDA</td>
<td>75.8 (+2.1)</td>
<td>78.7 (+2.4)</td>
</tr>
<tr>
<td>BERT\textsubscript{large}</td>
<td>78.7</td>
<td>81.9</td>
</tr>
<tr>
<td>+ CRQDA</td>
<td>80.6 (+1.9)</td>
<td>83.3 (+1.4)</td>
</tr>
<tr>
<td>RoBERT\textsubscript{a base}</td>
<td>78.6</td>
<td>81.6</td>
</tr>
<tr>
<td>+ CRQDA</td>
<td>80.2 (+1.6)</td>
<td>83.1 (+1.5)</td>
</tr>
<tr>
<td>RoBERT\textsubscript{a large}</td>
<td>86.0</td>
<td>88.9</td>
</tr>
<tr>
<td>+ CRQDA</td>
<td>86.4 (+0.4)</td>
<td>89.5 (+0.6)</td>
</tr>
</tbody>
</table>
Thanks