Erratum to Encoders Help You Disambiguate Word Senses in Neural Machine Translation

Gongbo Tang1 Rico Sennrich2,3 Joakim Nivre1
1Department of Linguistics and Philology, Uppsala University
2Institute of Computational Linguistics, University of Zurich
3School of Informatics, University of Edinburgh
firstname.lastname@{lingfil.uu.se, ed.ac.uk}

Abstract

Tang et al. (2019) explored the ability of NMT encoders and decoders to disambiguate word senses by evaluating hidden states. The work contained an error when evaluating decoder hidden states due to an implementation bug in generating decoder hidden states, which was discovered after the publication. After correcting the error, decoder hidden states from both Transformer and RNNS2S models achieve higher accuracy than encoder hidden states, which accords with the hypothesis that decoders provide further relevant information for disambiguation. The error does not otherwise affect the originally reported conclusions.

1 Error

Tang et al. (2019) explored the ability of NMT encoders and decoders to disambiguate word senses by evaluating hidden states. The work contained an error when evaluating decoder hidden states due to an implementation bug in generating decoder hidden states, which was discovered after the publication.

As a result of this error, the accuracy of decoder hidden states (DEC) reported in Table 2 does not reveal the ability of decoders in disambiguation. We have re-run the experiments with fixed decoder hidden states. A list of corrections are given below.

2 List of Corrections

- Abstract, the penultimate sentence:
  In contrast to encoders, the effect of decoder is different in models with different architectures.
  should become
  Decoders could provide further relevant information for disambiguation.

- Section 1, the third finding:
  Decoders hidden states have different effects on WSD in Transformer and RNNS2S.
  should be removed.

- Table 2:

<table>
<thead>
<tr>
<th></th>
<th>DE→EN</th>
<th></th>
<th>DE→FR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RNN. Trans.</td>
<td></td>
<td>RNN. Trans.</td>
<td></td>
</tr>
<tr>
<td>Embedding</td>
<td>63.1</td>
<td>63.2</td>
<td>68.7</td>
<td>68.9</td>
</tr>
<tr>
<td>ENC</td>
<td>94.2</td>
<td>97.2</td>
<td>91.7</td>
<td>95.6</td>
</tr>
<tr>
<td>DEC</td>
<td>97.9</td>
<td>91.2</td>
<td>95.1</td>
<td>91.6</td>
</tr>
</tbody>
</table>

- In Section 3.1, the last sentence:
  In addition, DEC achieves even higher accuracy than ENC in RNNS2S models but not in Transformer models.
  should be
  In addition, DEC achieves even higher accuracy than ENC in both RNNS2S models and Transformer models.

- Section 4.1.3:
  As Table 2 shows, RNN decoder hidden states could further improve the classification accuracy which accords with our hypothesis. It implies that the relevant information for WSD in the target-side has
been well incorporated into the decoder hidden states to predict the translations of ambiguous nouns. It is curious that Transformer decoder hidden states are inferior to Transformer encoder hidden states in our WSD classification task, given that Tang et al. (2018) and Rios et al. (2018) report better results with contrastive evaluation and semi-automatic evaluation of 1-best translations for Transformer models than for RNNS2S. However, note that our evaluation merely tests whether the information necessary for word sense disambiguation is encoded in hidden states and can be extracted by our binary classifier. In practice, decoder hidden states are used for predicting a target word from the entire vocabulary, and thus need to encode additional information which may confound our classifier. Despite these differences between RNNS2S and the Transformer, our results show that WSD is already possible on the basis of the encoder representation of the ambiguous noun, and that extracting contextual information via encoder-decoder attention or from the target history is not essential for WSD.

As Table 2 shows, decoder hidden states could further improve the classification accuracy which accords with our hypothesis. It implies that the relevant information for WSD in the target-side has been well incorporated into the decoder hidden states to predict the translations of ambiguous nouns. Although decoder hidden states achieve higher accuracy than encoder hidden states, the improvement is not as big as that achieved by encoder hidden states over word embeddings. This indicates that most of the disambiguation work is done by encoders.

- In Section 5, the last sentence of the first paragraph:
  Moreover, the effect of decoder hidden states on WSD is different in Transformer and RNNS2S models.
  should be
  Even though decoders could provide more relevant information for disambiguation, most of the disambiguation work is done by encoders.

- Footnote 6 in the Appendix:
  The classifiers fed decoder states are trained 200 epochs to converge.
  should be removed.

3 Conclusion

There was an implementation bug in generating decoder hidden states in Tang et al. (2019) and it was discovered subsequent to publication. After correcting the error, decoder hidden states from both Transformer and RNNS2S models achieve higher accuracy than those encoder hidden states. The error does not affect the originally reported statement that the encoder is the primary component for disambiguation.

References

