A Rating Task

A.1 Rating Instructions

Participants for the 5-star rating task were given the following instructions: “You will be presented with a German statement and a translation of this statement in English. You must assign a rating from 1 (Very Bad) to 5 (Very Good) to each translation.”

Participants for the pairwise task were given the following instructions: “You will be presented with a German statement and two translations of this statement in English. You must decide which of the two translations you prefer, or whether you have no preference.”

A.2 Example Ratings

Table 1 lists low- and high-variance items for 5-star ratings, Table 2 for pairwise ratings. From the annotations in the tables, the reader may get an impression which translations are “easier” to judge than others.

B Reward Estimation

B.1 Auxiliary Data for Reward Estimation

In order to augment the small collection of 1,000 rated translations, we leverage the available out-of-domain bitext as auxiliary training data: 10k source sentences of WMT (out-of-domain) are translated by the out-of-domain model. Translations from 9 beam search ranks are compared to their references to compute sBLEU rewards. This auxiliary data hence provides 90k out-of-domain training samples with sBLEU reward. For pairwise rewards, sBLEU scores for two translations for the same source are compared. Each mini-batch during training is sampled from the auxiliary data with probability $p_{\text{aux}}$, from the original training data with probability $1 - p_{\text{aux}}$. Adding this auxiliary data as a regularization through multi-task learning prevents the model from overfitting to the small set of human ratings. In our experiments, $p_{\text{aux}} = 0.8$ worked best.

B.2 Reward Estimation Architecture

Input source and target sequence are split into the BPE subwords used for NMT training, padded up to a maximum length of 100 tokens, and represented as 500-dimensional subword embeddings. Subword embeddings are pre-trained on the WMT bitext with word2vec (Mikolov et al., 2013), normalized to unit length and held constant during further training. Additional 10-dimensional BPE-feature embeddings are appended to the subword embeddings, where a binary indicator encodes whether each subword contains the subword prefix marker “@@”. BPE-prefix features are useful information for the model since bad translations can arise from “illegal” compositions of subword tokens. The embeddings are then fed to a source-side and a target-side bidirectional LSTM (biLSTM) (Hochreiter and Schmidhuber, 1997), respectively. The biLSTM outputs are concatenated for each time step and fed to a 1-D convolutional layer with 50 filters each for filter sizes from 2 to 15. The convolution is followed by max-over-time pooling, producing 700 input features for a fully-connected output layer with leaky ReLU (Maas et al., 2013) activation function. Dropout (Srivastava et al., 2014) with $p = 0.5$ is applied before the final layer. This architecture can be seen as a biLSTM-enhanced bilingual extension to the convolutional model for sentence classification proposed by Kim (2014).
These could be courier services, or technicians like, for example, just to make sure that the AED is still in its place.

It must be for me here and now, otherwise I cannot convince my audience that my concern is.

Table 1: Items with lowest (top) and highest (bottom) deviation in 5-star ratings. Mean normalized rating and standard deviation are reported. Problematic parts of source and target are underlined, namely hallucinated or inadequate target words (1, 5, 6), over-literal translations (2), ungrammatical source (3, 6) and omissions (4).

<table>
<thead>
<tr>
<th>#1</th>
<th>source</th>
<th>target</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zu diesem Zeitpunkt haben wir mehrzeitige Gemeinschaften, Gemeinschaften von vielen verschiedenen Zelltypen, welche zusammen als einher Organismus fungieren.</td>
<td>Es muss für mich im Hier und Jetzt stimmen sein, sonst kann ich mein Publikum nicht davon überzeugen, dass das mein Anliegen ist.</td>
<td>0.46, $\sigma = -0.30$</td>
</tr>
<tr>
<td></td>
<td>Wir fingen dann an, über Musik zu sprechen, angefangen von Bach über Beethoven, Brahms, Bruckner und all die anderen Bs, von Bartók bis hin zu Esa-Pekka Salonen.</td>
<td>It must be for me here and now, otherwise I cannot convince my audience that my concern is.</td>
<td>0.48, $\sigma = -0.70$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#2</th>
<th>source</th>
<th>target</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Finden Sie heraus, wie Sie überleben würden. Die meisten unserer Spieler haben die im Spiel gelernten Gewohnheiten beibehalten.</td>
<td>aber wenn Sie biologischen Evolution akzeptieren, bedenken Sie folgendes: ist es nur über die Vergangenheit, oder geht es auch um die Zukunft?</td>
<td>0.48, $\sigma = 1.12$</td>
</tr>
<tr>
<td></td>
<td>Finden Sie heraus, wie Sie überleben würden. Die meisten unserer Spieler haben die im Spiel gelernten Gewohnheiten beibehalten.</td>
<td>Aber wenn Sie biologischen Evolution akzeptieren, bedenken Sie folgendes: ist es nur über die Vergangenheit, oder geht es auch um die Zukunft?</td>
<td>0.48, $\sigma = -0.70$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#3</th>
<th>source</th>
<th>target</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nu können die googlen, aber es ist keine Infektion des Raehens sondern der oberen Atemwege und verursacht den Verschluss der Atemwege.</td>
<td>In this case we have multi-cell communities, communities of many different cell types, which act together as individual organism.</td>
<td>0.31, $\sigma = -0.52$</td>
</tr>
<tr>
<td></td>
<td>Nu können die googlen, aber es ist keine Infektion des Raehens sondern der oberen Atemwege und verursacht den Verschluss der Atemwege.</td>
<td>In this case we have multi-cell communities, communities of many different cell types, which act together as individual organism.</td>
<td>0.31, $\sigma = -0.52$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#4</th>
<th>source</th>
<th>target</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ich hielt meinen üblichen Vortrag, und danach sah sie mich an und sagte: &quot;Mhmm. Mhmm. Mhmm.&quot;</td>
<td>Wir durchgehen dieselben Stufen, welche Mehrzellerorganismen durchgemacht haben – Die Abstraktion unserer Methoden, wie wir Daten festhalten, präsentieren, verarbeiten.</td>
<td>1.25, $\sigma = -0.93$</td>
</tr>
<tr>
<td></td>
<td>Ich hielt meinen üblichen Vortrag, und danach sah sie mich an und sagte: &quot;Mhmm. Mhmm. Mhmm.&quot;</td>
<td>Wir durchgehen dieselben Stufen, welche Mehrzellerorganismen durchgemacht haben – Die Abstraktion unserer Methoden, wie wir Daten festhalten, präsentieren, verarbeiten.</td>
<td>1.25, $\sigma = -0.93$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#5</th>
<th>source</th>
<th>target</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>So in these Plans, we had about 657 plans that offered people anything between two to 59 different funds.</td>
<td>We pass the same steps that have passed through multi-cell organisms to process the abstraction of our methods, how we record data.</td>
<td>0.0, $\sigma = 1.0$</td>
</tr>
<tr>
<td></td>
<td>So in these Plans, we had about 657 plans that offered people anything between two to 59 different funds.</td>
<td>We pass the same steps that have passed through multi-cell organisms to process the abstraction of our methods, how we record data.</td>
<td>0.0, $\sigma = 1.0$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#6</th>
<th>source</th>
<th>target</th>
<th>rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>We went through the same steps that multicellular organisms have gone through – the abstraction of our methods of holding, representing, processing.</td>
<td>We go through the same steps that multicellular organisms have gone through – the abstraction of our methods of holding, representing, processing.</td>
<td>0.0, $\sigma = 1.0$</td>
</tr>
<tr>
<td></td>
<td>Well, it seems to me that this issue is going to be, or should be at least the most interesting political debate to follow.</td>
<td>We go through the same steps that multicellular organisms have gone through – the abstraction of our methods of holding, representing, processing.</td>
<td>0.0, $\sigma = 1.0$</td>
</tr>
</tbody>
</table>

Table 2: Items with lowest (top) and highest (bottom) deviation in pairwise ratings. Preferences of target1 are treated as ”−1“-ratings, preferences of target2 as ”1“, no preference as ”0“, so that a mean ratings of e.g. -0.14 expresses a slight preference of target1. Problematic parts of source and targets are underlined, namely hallucinated or inadequate target words (1, 2, 3, 4), correct target logic (2), omissions (3), ungrammatical source (4), capitalization (5), over-literal translations (5, 6).
C NMT

C.1 NMT Hyperparameters

The NMT has a bidirectional encoder and a single-layer decoder with 1,024 GRUs each, and subword embeddings of size 500 for a shared vocabulary of subwords obtained from 30k byte-pair merges (Sennrich et al., 2016). Maximum input and output sequence length are set to 60. For the MLE training of the out-of-domain model, we optimize the parameters with Adam ($\alpha = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$) (Kingma and Ba, 2014). For further in-domain tuning (supervised, OPL and RL), $\alpha$ is reduced to $10^{-5}$. To prevent the models from overfitting, dropout with probability 0.2 (Srivastava et al., 2014) and l2-regularization with weight $10^{-8}$ are applied during training. The gradient is clipped to its norm when its norm exceeds 1.0 (Pascanu et al., 2013). Early stopping points are determined on the respective development sets. For model selection we use greedy decoding, for test set evaluation beam search with a beam of width 10. For MLE and OPL models, mini-batches of size 60 are used. For the RL models, we reduce the batch size to 20 to fit $k = 5$ samples for each source into memory. The temperature is furthermore set to $\tau = 0.5$. We found that learning rate and temperature were the most critical hyperparameters and tuned both on the development set.

References


Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In ICML Workshop on Deep Learning for Audio, Speech and Language Processing. Atlanta, GA, USA.


