This results in:

- significantly higher accuracies than maximum-likelihood baseline
- better ranking of translation options, small
- but significant BLEU gains in English-to-

Our approach enables:

- accurate prediction of target translation
- stem and suffix given fixed amount of
- context
- automatic learning of relevant features with
- neural network architecture

This results in:

- significantly higher accuracies than
- maximum-likelihood baseline
- better ranking of translation options, small
- but significant BLEU gains in English-to-

Approach: Bilingual Neural Network (BNN)

Factorize word translation probability into stem and suffix probabilities: \[ p(t_j | c_s) = p(\sigma_j | c_s) p(\mu_j | c_s, \sigma_j) \]

Conditional probability normalized over the set of translation candidates instead of the whole output vocabulary:

\[ p(t_j | c_s) = \frac{\exp(W_2^{\sigma_j} \phi(W_1 r(c_{s-j})))}{\sum_{t_k \in \text{GEN}(c_{s-j})} \exp(W_2^{\sigma_k} \phi(W_1 r(c_{s-j})))} \]

Input: fixed-size source context window

Translation prediction results

Bilingual neural networks (BNN) prediction accuracy compared to a context-independent maximum likelihood baseline:

<table>
<thead>
<tr>
<th>SMT system</th>
<th>wmt12 (dev)</th>
<th>wmt13 (test)</th>
<th>Reference/MT-search-space [top-3%]</th>
<th>Reference/MT-search-space [top-30%]</th>
<th>Reference/MT-output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>24.7</td>
<td>18.9</td>
<td>57.6%</td>
<td>59.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>+ BNN</td>
<td>24.5</td>
<td>19.2</td>
<td>70.7%</td>
<td>70.9%</td>
<td>50.7%</td>
</tr>
<tr>
<td>+ stem/suff. BNN</td>
<td>24.5</td>
<td>19.2</td>
<td>85.0%</td>
<td>85.0%</td>
<td></td>
</tr>
<tr>
<td>+ word. BNN</td>
<td>24.7</td>
<td>19.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ stem/suff. BNN</td>
<td>24.7</td>
<td>19.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SMT results

Compute BNN score for each phrase pair, similarly to lexical weighting:

\[ p_{\text{BNN}}(\delta, t, a) = \prod_{i=1}^{n} \left( \frac{1}{|Q_a|} \sum_{j \in Q_a} \frac{P_{\text{BNN}}(t_j, c_s)}{P_{\text{BNN}}(\text{NULL}, c_s)} \right) \]

Effect of our BNN models on English-to-Russian translation quality (BLEU%):

- MLE
- BNN (p) stem suffix
- 0.6 0.6 0.7
- 0.1 0.6 0.1

Target word coverage analysis of the English-to-Russian SMT system before and after adding the morphological BNN models:

- Reference/MT-search-space [top-1] 57.6% 59.0%
- Reference/MT-search-space [top-3] 70.7% 70.9%
- Reference/MT-search-space [top-30] 85.0% 85.0%
- Reference/MT-output 50.0% 50.7%