A Neural Approach to Pun Generation
Zhiwei Yu, Jiwei Tan and Xiaojun Wan
Institute of Computer Science and Technology, Peking University
The MOE Key Laboratory of Computational Linguistics, Peking University
{yuzw,tanjiwei,wanxiaojun}@pku.edu.cn

Pun
A pun is a means of expression, the essence of which is in the given context the word or phrase can be understood in two meanings simultaneously.

Joint Model
Taking “count” for example, “countv01” means “determine the number or amount of”, while “countv08” means “have faith or confidence in”.

Figure 1: Framework of the proposed Joint Model

Conditional Language Model
Input: A pseudo-word (e.g. countv01).
Backward Model: to generate the backward sequence starting from the assigned pseudo-word and ending up with “</s>” (e.g. countv01 couldn’t mathematician inept the <s> ).
Forward Model: reverse the backward sequence as input to generate the forward sequence.
Output: A sentence contains the assigned pseudo-word.

Joint Beam Search Algorithm
Provided with two pseudo-words as inputs to the encoder in the backward generation process (e.g. “countv01” as input1 and “countv08” as input2), we decode two output sentences in parallel and the two sentences should be the same except for the input pseudo-words.
Our joint beam search algorithm selects candidates while decoding for the two inputs according to the joint score distribution on all beams.
The decoding process will be finished after all the beams have selected “</s>”.

Highlight Model
We improve the pun generation model by adding some associative words to the sentence which could remind people some special sense of the target word.

Word Association
Increase the probability of the associative words to be chosen according to their PMI scores.

Multinomial Sampling
Sampling is useful in cases where we may want to get a variety of outputs for a particular input, we use multinomial sampling to increase the uncertainty when generating the pun.

Evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>PPL</th>
<th>d. -1(%)</th>
<th>d. -2(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highlight</td>
<td>91.80</td>
<td>27.13</td>
<td>62.85</td>
</tr>
<tr>
<td>Joint</td>
<td>63.48</td>
<td>22.13</td>
<td>50.59</td>
</tr>
<tr>
<td>Normal Language</td>
<td>62.66</td>
<td>19.60</td>
<td>41.62</td>
</tr>
<tr>
<td>Pun Language</td>
<td>889.07</td>
<td>14.78</td>
<td>23.11</td>
</tr>
</tbody>
</table>

Table 1: Results of automatic evaluation

Example
square: 1) a plane rectangle with four equal sides and four right angles, a four-sided regular polygon; 2) someone who doesn’t understand what is going on .
Highlight: little is known when he goes back to the square of the football club
Joint: there is a square of the family
Normal Language: the population density was # people per square mile
Pun Language: when the pirate captain’s ship ran aground he couldn’t fathom why
Gold Puns: my advanced geometry class is full of squares

Conclusion
We proposed two models for pun generation without using training data of puns. Joint Model makes use of conditional language model and the joint beam search algorithm, which can assure the assigned senses of target words suitable in one sentence. Highlight Model takes associative words into consideration, which makes the distinct senses more obvious in one sentence.