

1. 用 Pointer-generator network 可以从原文中复制单词，从而生成字典中不存在的单词。其实现是通过设置了一个概率 p ，将通过 attention 生成的字典上的词概率乘以 p ，加上对应原文中单词的 attention 乘以 $(1-p)$ ，得到一个新的概率分布，这个分布是在字典和原文中的词上的，因此能够实现从原文中摘抄字典中不存在的词。

2. 引入 Coverage loss 可以有效地减少模型生成重复的话。其实现是通过记录之前的 attention 值，并输入进每一步的输出，从而让模型知道以前输出过哪些信息，另外在 loss 里增加 Coverage loss。