学习汇报

焦强

实验

基础的Seq2Seq，输入一个单词（字母序列），模型将返回一个对字母排序后的“单词”。

```
def seq2seq_model(input_data, targets, lr, target_sequence_length,
                  max_target_sequence_length, source_sequence_length,
                  source_vocab_size, target_vocab_size,
                  encoder_embedding_size, decoder_embedding_size,
                  rnn_size, num_layers):

    # 获取encoder的状态输出
    _, encoder_state = get_encoder_layer(input_data,
                                         rnn_size,
                                         num_layers,
                                         source_sequence_length,
                                         source_vocab_size,
                                         encoding_embedding_size)


    # 预处理后的decoder输入
    decoder_input = process_decoder_input(targets, target_letter_to_int, batch_size)

    # 将状态向量与输入传递给decoder
    training_decoder_output, predicting_decoder_output = decoding_layer(target_letter_to_int,
                                                                        decoding_embedding_size,
                                                                        num_layers,
                                                                        rnn_size,
                                                                        target_sequence_length,
                                                                        max_target_sequence_length,
                                                                        encoder_state,
                                                                        decoder_input)


    return training_decoder_output, predicting_decoder_output
```

```
Epoch  36/60 Batch   50/77 - Training Loss:  0.057  - Validation loss:  0.057
Epoch  37/60 Batch   50/77 - Training Loss:  0.052  - Validation loss:  0.053
Epoch  38/60 Batch   50/77 - Training Loss:  0.047  - Validation loss:  0.049
Epoch  39/60 Batch   50/77 - Training Loss:  0.043  - Validation loss:  0.045
Epoch  40/60 Batch   50/77 - Training Loss:  0.039  - Validation loss:  0.042
Epoch  41/60 Batch   50/77 - Training Loss:  0.036  - Validation loss:  0.039
Epoch  42/60 Batch   50/77 - Training Loss:  0.033  - Validation loss:  0.037
Epoch  43/60 Batch   50/77 - Training Loss:  0.030  - Validation loss:  0.034
Epoch  44/60 Batch   50/77 - Training Loss:  0.028  - Validation loss:  0.032
Epoch  45/60 Batch   50/77 - Training Loss:  0.026  - Validation loss:  0.029
Epoch  46/60 Batch   50/77 - Training Loss:  0.024  - Validation loss:  0.028
Epoch  47/60 Batch   50/77 - Training Loss:  0.027  - Validation loss:  0.029
Epoch  48/60 Batch   50/77 - Training Loss:  0.030  - Validation loss:  0.030
Epoch  49/60 Batch   50/77 - Training Loss:  0.023  - Validation loss:  0.026
Epoch  50/60 Batch   50/77 - Training Loss:  0.021  - Validation loss:  0.024
Epoch  51/60 Batch   50/77 - Training Loss:  0.019  - Validation loss:  0.022
Epoch  52/60 Batch   50/77 - Training Loss:  0.017  - Validation loss:  0.021
Epoch  53/60 Batch   50/77 - Training Loss:  0.016  - Validation loss:  0.020
Epoch  54/60 Batch   50/77 - Training Loss:  0.015  - Validation loss:  0.019
Epoch  55/60 Batch   50/77 - Training Loss:  0.014  - Validation loss:  0.018
Epoch  56/60 Batch   50/77 - Training Loss:  0.013  - Validation loss:  0.018
Epoch  57/60 Batch   50/77 - Training Loss:  0.012  - Validation loss:  0.017
Epoch  58/60 Batch   50/77 - Training Loss:  0.011  - Validation loss:  0.016
Epoch  59/60 Batch   50/77 - Training Loss:  0.011  - Validation loss:  0.016
Epoch  60/60 Batch   50/77 - Training Loss:  0.010  - Validation loss:  0.015
Model Trained and Saved
```

```
原始输入: common

Source
  Word 编号:    [20, 28, 6, 6, 28, 5, 0]
  Input Words: c o m m o n <PAD>

Target
  Word 编号:       [20, 6, 6, 5, 28, 28, 3]
  Response Words: c m m n o o <EOS>
```

论文

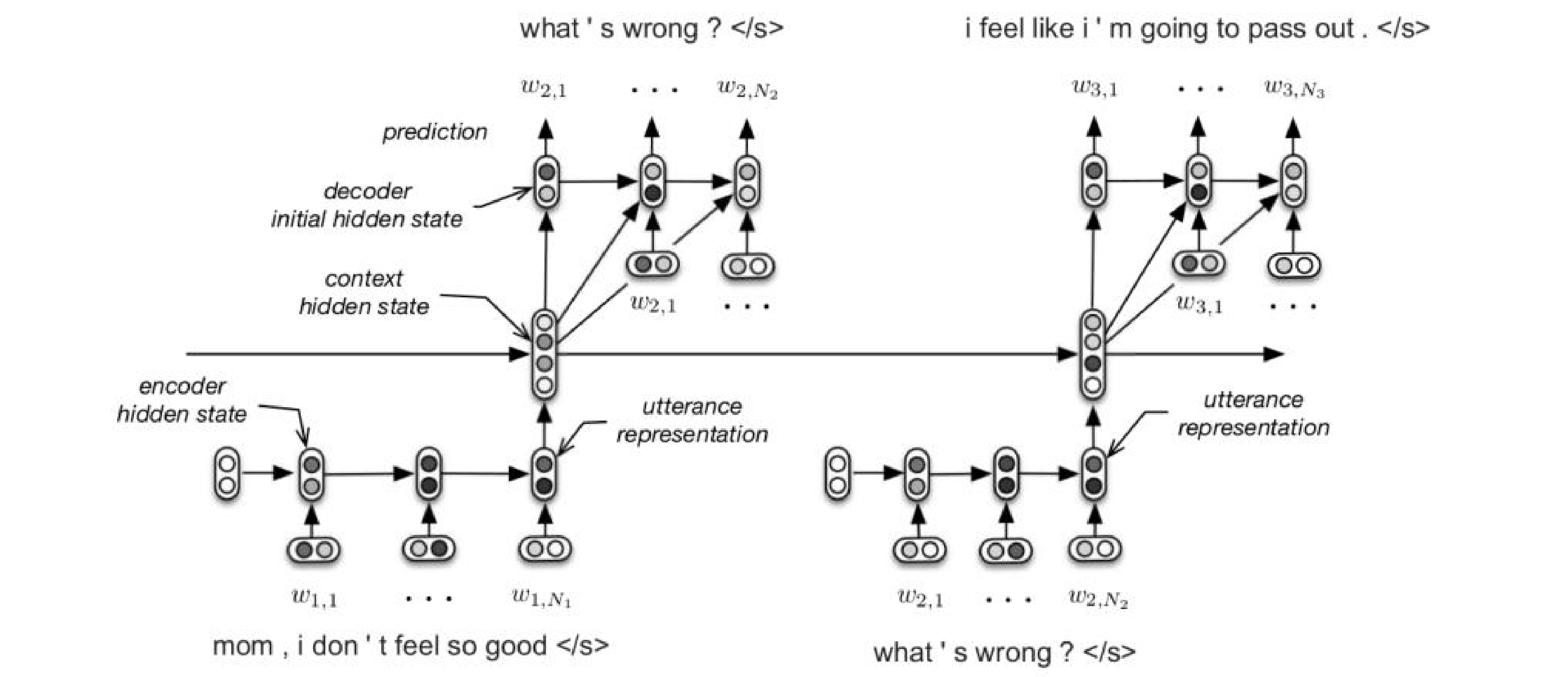
Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models

AAAI 2016

https://xueshu.baidu.com/usercenter/paper/show?paperid=d2ffcde5529ecf0ce24db265fd9ce3dc&site=xueshu_se&hitarticle=1

将分层的递归编码器-解码器神经网络扩展到对话领域，采用分层的seq2seq架构来解决多轮对话问题。

模型：



结果：

Model	Perplexity	Perplexity@U_3	Error-Rate	Error-Rate@U_3
Backoff N-Gram	64.89	65.05	-	-
Modified Kneser-Ney	60.11	54.75	-	-
Absolute Discounting N-Gram	56.98	57.06	-	-
Witten-Bell Discounting N-Gram	53.30	53.34	-	-
RNN	35.63 ± 0.16	35.30 ± 0.22	$66.34\% \pm 0.06$	$66.32\% \pm 0.08$
DCGM-I	36.10 ± 0.17	36.14 ± 0.26	$66.44\% \pm 0.06$	$66.57\% \pm 0.10$
HRED	36.59 ± 0.19	36.26 ± 0.29	$66.32\% \pm 0.06$	$66.32\% \pm 0.11$
HRED + Word2Vec	33.95 ± 0.16	33.62 ± 0.25	$66.06\% \pm 0.06$	$66.05\% \pm 0.09$
RNN + SubTle	27.09 ± 0.13	26.67 ± 0.19	$64.10\% \pm 0.06$	$64.07\% \pm 0.10$
HRED + SubTle	27.14 ± 0.12	26.60 ± 0.19	$64.10\% \pm 0.06$	$64.03\% \pm 0.10$
HRED-Bi. + SubTle	$\mathbf{26.81 \pm 0.11}$	$\mathbf{26.31 \pm 0.19}$	$\mathbf{63.93\% \pm 0.06}$	$\mathbf{63.91\% \pm 0.09}$

Table 2: Test set results computed on $\{U_1, U_2, U_3\}$ and solely on $\{U_3\}$ conditioned on $\{U_1, U_2\}$. Standard deviations are shown for all neural models. Best performances are marked in bold.

Reference (U_1, U_2)	**MAP**	**Target** (U_3)
U_1: yeah , okay . U_2: well , i guess i ' ll be going now .	i ' ll see you tomorrow .	yeah .
U_1: oh . <continued_utterance> oh . U_2: what ' s the matter , honey ?	i don ' t know .	oh .
U_1: it ' s the cheapest . U_2: then it ' s the worst kind ?	no , it ' s not .	they ' re all good , sir .
U_1: <person> ! what are you doing ? U_2: shut up ! c ' mon .	what are you doing here ?	what are you that crazy ?

Table 3: MAP outputs for HRED-Bidirectional bootstrapped from *SubTle* corpus. The first column shows the reference utterances, where U_1 and U_2 are respectively the first and second utterance in the test triple. The second column shows the MAP output produced by beam-search conditioned on U_1 and U_2. The third column shows the actual third utterance in the test triple.

谢谢