



## 基于双编码器结构的文本自动摘要研究

冯读娟, 杨 璐, 严建峰

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

**摘 要:** 为了解决序列到序列模型中编码器不能充分编码源文本的问题, 构建一种基于双编码器网络结构的 CGAtten-GRU 模型。2 个编码器分别使用卷积神经网络和双向门控循环单元, 源文本并行进入双编码器, 结合 2 种编码网络结构的输出结果构建注意力机制, 解码器端使用 GRU 网络融合 Copy 机制和集束搜索方法, 以提高解码的准确度。在大规模中文短文本摘要数据集 LCSTS 上的实验结果表明, 与 RNN context 模型相比, 该模型的 Rouge-1、Rouge-2 和 Rouge-L 分别提高 0.1、0.059 和 0.046。

**关键词:** 自然语言处理; 生成式摘要; 卷积神经网络; 门控循环单元; 注意力机制; 序列到序列模型; Copy 机制

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 冯读娟, 杨璐, 严建峰. 基于双编码器结构的文本自动摘要研究[J]. 计算机工程, 2020, 46(6): 60-64.

**英文引用格式:** FENG Dujuan, YANG Lu, YAN Jianfeng. Research on automatic text summarization based on dual-encoder structure[J]. Computer Engineering, 2020, 46(6): 60-64.

## Research on Automatic Text Summarization Based on Dual-Encoder Structure

FENG Dujuan, YANG Lu, YAN Jianfeng

(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

**[Abstract]** This paper constructs a CGAtten-GRU model based on dual-encoder network structure to solve the problem that the encoder cannot fully encode the source text in the sequence-to-sequence(seq2seq) model. The two encoders use Convolutional Neural Network(CNN) and Bidirectional Gated Recurrent Unit(BiGRU) respectively, and the source text enters the two encoders in parallel. An attention mechanism is constructed by means of the outputs of two encoding networks. The decoder uses GRU network combining the Copy mechanism and the beam search method to improve the accuracy of decoding. Experimental results on large-scale Chinese short text summarization dataset LCSTS show that compared with the RNN context model, the proposed model improves Rouge-1 by 0.1, Rouge-2 by 0.059, and Rouge-L by 0.046.

**[Key words]** Natural Language Processing(NLP); abstractive summarization; Convolutional Neural Network(CNN); Gated Recurrent Unit(GRU); attention mechanism; sequence-to-sequence(seq2seq) model; Copy mechanism

**DOI:** 10.19678/j.issn.1000-3428.0054540

### 0 概述

文本摘要的目的是简化句子以抽取文本主要信息。目前, 文本摘要按照文档数量可以分为单文档摘要和多文档摘要, 按照实现方式可以分为抽取式摘要和生成式摘要。抽取式自动摘要按照一定的权重计算方法抽取原文本中的重要句子并进行拼接以组成文本摘要<sup>[1]</sup>。生成式摘要则是计算机读取原文本后, 在理解整篇文章的基础上不使用原文句子而直接生成的摘要结果。随着深度学习技术在自然语言

处理(Natural Language Processing, NLP)领域的广泛应用, 特别是序列到序列(sequence-to-sequence, seq2seq)模型的出现, 使得生成式摘要更加流畅和自然。

来自 Facebook 的 RUSH 等人<sup>[2]</sup>较早提出一种 Encoder-Decoder 框架下的句子摘要模型, 其中尝试 Bag-of-Words、CNN 和 Attention-Based 3 种编码方式, 同时还与抽取式方法相结合, 分别在 DUC2004 和 Gigaword 数据集上进行对比, 结果表明, 增加了人工特征的 ABS+ 模型效果优于 ABS 模型。

CHOPRA 等人<sup>[3]</sup>同样使用卷积网络编码原文,

**基金项目:** 国家自然科学基金(61572339, 61272449); 江苏省科技支撑计划重点项目(BE2014005)。

**作者简介:** 冯读娟(1994—), 女, 硕士研究生, 主研方向为自然语言处理、自动摘要; 杨 璐、严建峰, 副教授。

**收稿日期:** 2019-04-09 **修回日期:** 2019-06-11 **E-mail:** fdujuan29@163.com

使用循环神经网络(RNN)作为解码器。根据 RNN 作用的不同,将模型分为 RAS-Elman 和 RAS-LSTM, 在 Gigaword 数据集上对比各模型,结果表明,RAS-Elman 模型效果较好。

来自 IBM 的 NALLAPATI 等人<sup>[4]</sup>使用双向门控循环单元(BiGRU)作为编码器对原文实现编码,解码器使用 GRU 进行解码,同时融合一些特征到模型中,实验结果表明,该方法的性能优于单一使用深度神经网络。BAHDANAU 等人<sup>[5]</sup>提出一种新的解码方式,其解码的源头不仅是编码输出的语义向量,同时构建一种从输入序列中为当前预测词自动搜寻相关部分的机制,即注意力机制,该机制有效提高了神经机器翻译(Neural Machine Translation, NMT)的效果。SEE 等人<sup>[6]</sup>采用 BiLSTM 编码原文的词向量序列,解码器采用单向 LSTM,使用指针网络模型缓解未登录词(Out-of-Vocabulary, OOV)问题,参数复用降低了模型的复杂度。GEHRING 等人<sup>[7]</sup>将 CNN 用于神经机器翻译任务上,其相比 BiLSTM 取得了较好的 BLEU 结果,且训练速度得到较大提高。

本文建立一种新的 CGAtten-GRU 双编码网络模型以解决文本自动摘要生成问题。通过多层 CNN 和 BiGRU 分别构建 2 个不同的编码器,利用双编码器输出结果建立注意力机制。解码器端使用 GRU 同时融合 Copy 机制和集束搜索(Beam-Search)方法进行解码。在大规模中文短文本摘要数据集 LCSTS 上进行实验,将 Rouge<sup>[8]</sup>作为评价指标,以验证该模型的性能。

## 1 基于双编码器结构的生成式摘要模型

在自然语言处理过程中,seq2seq 模型普遍使用 RNN 作为编码器和解码器。编码器 RNN 处理长度为  $m$  的输入序列  $x = (x_1, x_2, \dots, x_m)$  并返回状态向量  $h = (h_1, h_2, \dots, h_m)$ 。中间向量  $h$  进入编码器 RNN 生成结果  $y = (y_1, y_2, \dots, y_n)$ 。解码端利用上一个解码状态  $s_i$ 、前一个目标词汇的 embedding 向量  $w_i$  以及由编码输出构造的语义向量  $c_i$  共同生成下一个时间步的输出  $y_{i+1}$ 。不同的 RNN 种类以及不同的条件输入,将构造出不同的编码器-解码器架构。

RNN 具有记忆功能,每个时间步状态  $s_i$  捕捉之前时间点上的信息。为了规避标准 RNN 中的梯度爆炸和梯度消失问题,一般使用 RNN 中的 2 种改进网络结构,分别是长短时记忆(Long Short-Term Memory, LSTM)网络和 GRU。LSTM<sup>[9]</sup>网络相对于 RNN 学习速度更快,而 GRU 网络<sup>[10]</sup>是 LSTM 网络的一种变体,其将 LSTM 的忘记门和输入门合成为一个单一的更新门,同时混合了细胞状态和隐藏状态等。本文模型中全部使用更为简单快速的 GRU 网络。

目前,在自然语言处理领域,CNN 在机器翻译、文本分类等任务中已经有许多应用<sup>[11-13]</sup>。卷积网络在输入序列的固定大小窗口上操作,使得能够同时计算源句子的所有特征。RNN 网络维持整个过去的隐藏状态,无法进行序列内的并行计算。与 RNN 相比,一系列卷积层操作为获取序列元素之间的关系提供了更快捷的方法。卷积网络处理过程自上而下,所有单词都经历相同数量的变换,而 RNN 的第一个单词被过度处理且最后一个单词仅被处理一次。

基于以上 2 种网络结构在编码器-解码器架构中的应用优势与缺点,本文尝试将 2 种网络结构相融合,构造一种双编码网络结构以对源文本进行充分编码。

将注意力机制引入到初始的编码器-解码器框架中是为了解决传统结构在编码解码时仅依赖于内部一个固定长度向量的问题<sup>[14]</sup>。单纯的编码器-解码器模型在编码时将所有输入信息压缩成固定长度传递给解码器,这个过程中存在信息丢失问题,导致解码器接收到的语义向量无法完全表示整个序列的信息。此外,输出词汇更多的是受输入序列某些特定部分的影响<sup>[15]</sup>,解码器在生成词汇时缺少选择性关注输入序列的机制,影响了输出的准确率。

注意力机制通过在输入序列上引入注意力权重,优先考虑某些相关信息的位置。本文提出的双编码器模型分别经过卷积网络和 RNN,通过结合双编码器的中间输出结果来构造新的上下文向量。

本文 CGAtten-GRU 模型的整体结构如图 1 所示。其中,一个编码器 encoder-1 使用 3 层卷积网络,另一个编码器 encoder-2 使用单层双向 GRU 网络,解码器使用单层 GRU 网络。

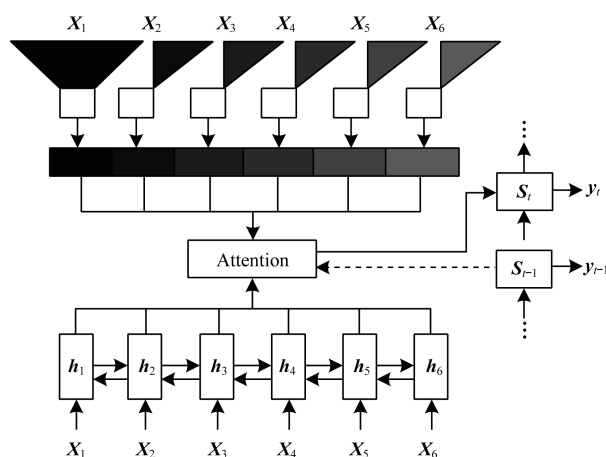


图 1 CGAtten-GRU 模型整体结构

Fig. 1 Overall structure of the CGAtten-GRU model

## 2 CGAtten-GRU 模型构建

### 2.1 编码器

编码器 encoder-1 采用单层 BiGRU 网络读取源文本。GRU 网络定义如下:

$$z_i = \sigma(W_z[x_i, h_{i-1}]) \quad (1)$$

$$r_i = \sigma(W_r[x_i, h_{i-1}]) \quad (2)$$

$$\tilde{h}_i = \tanh(W_h[x_i, r_i \odot h_{i-1}]) \quad (3)$$

$$h_i = (1 - z_i) \odot h_{i-1} + z_i \odot \tilde{h}_i \quad (4)$$

其中,  $W_z$ 、 $W_r$ 、 $W_h$  为权重矩阵。

双向 GRU 网络包含前向神经网络和后向神经网络。由于单向网络结构仅将已经输入的过去的信息存储在隐藏状态中, 训练过程中无法使用未来的信息, 因此本文实验中使用双向网络结构<sup>[16]</sup>, 同时保留来自正负 2 个时间方向的状态数据。

前向网络从左往右读取源文本的词嵌入向量得到一个隐藏状态序列  $(\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n)$ , 后向网络从右往左读取源文本的词嵌入向量得到另一个隐藏状态序列  $(\overleftarrow{h}_1, \overleftarrow{h}_2, \dots, \overleftarrow{h}_n)$ 。

$$\vec{h}_i = GRU(x_i, \vec{h}_{i-1}) \quad (5)$$

$$\overleftarrow{h}_i = GRU(x_i, \overleftarrow{h}_{i+1}) \quad (6)$$

BiGRU 初始状态设置为零向量。编码器 encoder-1 读取源文本后, 级联前向网络及后向网络的隐藏状态, 即  $h_i = [\vec{h}_i, \overleftarrow{h}_i]$ 。

编码器 encoder-2 由 3 层卷积网络堆叠而成。文献[17]研究表明, 实验过程中卷积网络设置越深的层数, 得出的实验结果就越好。但是达到一定深度后, 模型的性能不会随着网络层数的增加而一直提升。在本文的中文文本生成式摘要任务中, 设置卷积层数为 3 并作为 encoder-2 编码器, 此时能取得较好的实验结果。

在使用卷积网络时, 需要对固定长度的源文本进行填充, 在本文实验中, 步幅长度设为 1, 源文本句首及句尾分别填充 1 个词汇长度, 卷积核大小为  $(k, d)$ , 其中,  $d$  为词嵌入向量的维度,  $k$  大小设置为 3。以上设置能确保输出长度仍然等于输入长度,  $o$  为卷积网络输出向量, 即  $|x| = |o|^{[7]}$ 。

### 2.2 解码器

本文使用由双编码器结构计算出的注意力并结合 GRU 网络共同解码生成摘要。在每个解码时间步  $t$  时刻, GRU 读取前一个词向量  $w_{t-1}$  和前一个语义向量  $c_{t-1}$  以计算新的状态  $s_t$ 。其中, 解码的初始转态  $s_0$  由编码器 encoder-1 的最后一个后向隐藏状态  $\overleftarrow{h}_1$  计算得到, 即:

$$s_t = GRU(w_{t-1}, c_{t-1}, s_{t-1}) \quad (7)$$

$$s_0 = \tanh(W_d \overleftarrow{h}_1 + b) \quad (8)$$

其中,  $W_d$  是权重矩阵,  $b$  是偏置向量。

语义向量  $c_t$  的权重  $a_{t,i}$  由 encoder-2 卷积网络的输出结果计算:

$$e_{t,i} = v_a^T \tanh(W_a s_{t-1} + U_a o) \quad (9)$$

$$a_{t,i} = \frac{\exp(e_{t,i})}{\sum_{i=1}^n \exp(e_{t,i})} \quad (10)$$

对以上结果和 encoder-1 的 BiGRU 网络输出状态  $h_i$  进行加权求和,  $c_t$  的最终结果为:

$$c_t = \sum_{i=1}^n a_{t,i} h_i \quad (11)$$

结合前一个词的词嵌入向量  $w_{t-1}$ 、当前语义表示  $c_t$ 、解码状态  $s_t$  三者计算获得输出状态  $r_t$ 。 $r_t$  通过一个 maxout<sup>[18]</sup> 隐藏层与一个 softmax 层后得到一个词在词典中的位置:

$$r_t = W_r w_{t-1} + U_r c_t + V_r s_t \quad (12)$$

$$m_t = [\max\{r_{t,2j-1}, r_{t,2j}\}]_{j=1,2,\dots,d}^T \quad (13)$$

$$p(y_t | y_1, y_2, \dots, y_{t-1}) = \text{softmax}(W_o m_t) \quad (14)$$

### 2.3 Copy 机制与集束搜索

由于摘要结果往往包含一些命名实体和稀有名词, 且语言本身也会不断产生一些新词汇, 即使增大词典也不可能将所有实体名词和新词等全部包含在内, 这一问题对摘要结果影响较大。本文模型通过 Copy 机制<sup>[19-20]</sup> 及 Beam-Search<sup>[21]</sup> 集束搜索方法提高摘要准确性。

在模型中使用 Copy 机制时, 有 2 个 softmax 层来预测下一个词汇, 一个预测该词在原语句中的位置, 另一个预测该词在词典中的位置。从源文本复制单词的概率为:

$$p = \sigma(Ws_t + Uc_t + b) \quad (15)$$

集束搜索从左到右生成摘要词汇, 每次生成一个词汇时均对结果进行排序, 并保留固定数量的候选词汇, 固定的数量即集束宽度。集束搜索通过这种剪枝操作能够在一定程度上降低计算复杂度, 提高解码速度。本文模型设置集束宽度为 12。

## 3 实验结果与分析

### 3.1 实验数据集

本文实验使用 LCSTS 数据集<sup>[22]</sup>, 该数据集包含 3 个部分, 如表 1 所示。

1) 第 1 部分包含 2 400 591 个短文本-摘要对, 是数据集的主要构成部分, 这部分数据在实验中用作训练集。

2) 第 2 部分包含 10 666 个文本-摘要训练对, 全

部从第1部分数据中随机抽取,人工打分标注在1~5之间,表示文本与对应摘要之间的相关性,“1”表示最不相关,“5”表示最匹配。

3)第3部分包含1 106个文本-摘要对。对于这部分数据,3位志愿者分别独立地对2 000个文本数据进行标记,提取具有相同分数的数据得到1 106个文本-摘要对。本部分数据集独立于第1部分与第2部分。在实验中,本文使用该部分3、4、5评分的数据作为短文本摘要生成任务的测试集。

表1 数据集信息

Table 1 Dataset information

| 数据量及人工评分  | 第1部分      | 第2部分   | 第3部分  |
|-----------|-----------|--------|-------|
| 短文本-摘要对数量 | 2 400 591 | 10 666 | 1 106 |
| 人工评分1分    | —         | 942    | 165   |
| 人工评分2分    | —         | 1 039  | 216   |
| 人工评分3分    | —         | 2 019  | 227   |
| 人工评分4分    | —         | 3 128  | 301   |
| 人工评分5分    | —         | 3 538  | 197   |

短文本(ST)、分词短文本(Segmented ST)、摘要(SUM)和分词摘要(Segmented SUM)长度分布情况如图2所示<sup>[22]</sup>。

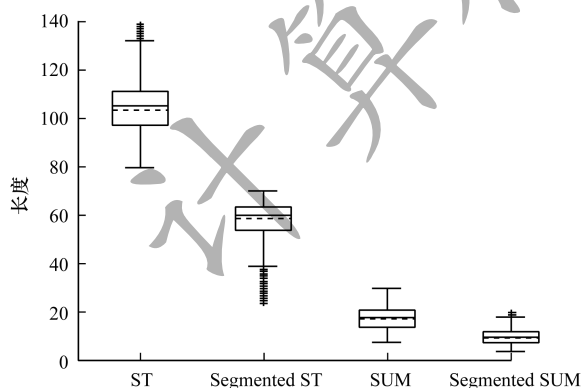


图2 数据集文本及摘要长度统计

Fig. 2 Dataset text and summary length statistics

### 3.2 实验设置

实验将第1部分数据作为训练集,通过jieba工具将文本及摘要进行分词处理。如图2所示,分词文本长度中位数略大于60,分词摘要长度中位数约为10,因此,在本文实验中,固定文本及摘要长度分别为65和15。剔除分词后文本长度小于10、摘要长度小于5的数据。实际训练数据量为2 340 205,过滤掉60 343条长度不符合要求的数据对。

固定词汇表大小为50 000,从分词处理的数据集中选取50 000个高频词作为词汇表。本文设计的解码器词汇表和编码器相同,同时共用编码器的词嵌入矩阵。批尺寸大小为64,词向量维度为256。本文采用Adam<sup>[23]</sup>模型优化器,学习率为0.001。所有模型的训练过程都在GTX 1080Ti机器上完成,实验用时2天左右。

测试集使用第3部分数据中评分为3、4、5的文本-摘要组合对,共计725个。实验结果使用Rouge<sup>[8]</sup>进行评价。Rouge是国际公认的自动摘要评价标准,其通过计算标准摘要与实验摘要结果重叠词汇来衡量摘要质量,例如,使用单词汇、双词汇以及最长公共子序列词汇长度分别计算Rouge-1、Rouge-2、Rouge-L值。在下文的实验过程中,主要使用Rouge-1、Rouge-2、Rouge-L 3个值评价本文的生成式摘要质量,并与文献[22]提出的2个模型进行实验结果对比。

### 3.3 结果分析

如表2所示,第1组实验模型结构(RNN)为文献[22]提出的第1个模型,其未引入注意力机制,使用RNN做编码器和解码器,且将编码器得到的最后一个隐藏层状态传输到解码器;第2组实验模型结构(RNN context)为文献[22]提出的第2个模型,其第1组模型的基础上在解码器端引入上下文,将编码器所有状态传入解码器。以上2组实验模型中的RNN结构均采用GRU网络,使用Adadelta作为模型优化器,解码过程中利用集束搜索得到摘要结果。第3组编码器(MC-GRU)仅采用CNN网络,卷积结构与本文模型相同,解码器使用RNN,包括注意力机制。第4组为本文模型,其采用2种编码器结构构建注意力机制,解码器端采用了GRU网络,同时引入注意力及Copy机制并使用集束搜索方法。从表2可以看出,在Rouge评价系统下,本文CGAtten-GRU模型的Rouge-1、Rouge-2、Rouge-L值高于第2组和第3组模型。CGAtten-GRU模型引入2个编码器构造注意力,有效弥补了编码后中间语义向量包含信息不充分的问题,提高了进入解码器的信息量,从而极大地提升了摘要效果。同时,本文模型在解码器端辅助使用Copy机制和集束搜索方法,也在一定程度上提高了摘要的准确性。

表2 4种模型性能对比

Table 2 Performance comparison of four models

| 指标      | RNN   | RNN context | MC-GRU | CGAtten-GRU |
|---------|-------|-------------|--------|-------------|
| Rouge-1 | 0.177 | 0.268       | 0.336  | 0.368       |
| Rouge-2 | 0.085 | 0.161       | 0.192  | 0.220       |
| Rouge-L | 0.158 | 0.241       | 0.245  | 0.287       |

## 4 结束语

本文提出一种基于双编码器结构的CGAtten-GRU模型。构造2种不同的编码器网络结构对源文本并行编码,2个编码器输出的语义向量共同构造注意力机制,以丰富编码输出的中间信息内容,使得解码获取足够的信息从而提高其准确度。结合Copy机制及集束搜索方法解决未登录词问题,进一步提升摘要效果。在大规模中文数据集LCSTS上的实验结果表明,该模型生成的中文摘要质量较高。下一步将改进卷积编码器结构并提高注意力区分度,以改善摘要结果中的词语重复现象。

## 参考文献

- [ 1 ] PENG Min, GAO Binlong, HUANG Jimin, et al. Automatic summarization of microblog based on high quality information extraction [ J ]. Computer Engineering, 2015, 41 ( 7 ) : 36-42. ( in Chinese )  
彭敏,高斌龙,黄济民,等.基于高质量信息提取的微博自动摘要[ J ]. 计算机工程, 2015, 41 ( 7 ) : 36-42.
- [ 2 ] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization [ EB/OL ]. [ 2019-03-20 ]. <https://arxiv.org/pdf/1509.00685.pdf>.
- [ 3 ] CHOPRA S, AULI M, RUSH A M. Abstractive sentence summarization with attentive recurrent neural networks [ EB/OL ]. [ 2019-03-20 ]. <http://aclweb.org/anthology/N/N16/N16-1012.pdf>.
- [ 4 ] NALLAPATI R, ZHOU B, GULCEHRE C, et al. Abstractive text summarization using sequence-to-sequence RNNs and beyond [ EB/OL ]. [ 2019-03-20 ]. <https://arxiv.org/pdf/1602.06023.pdf>.
- [ 5 ] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [ EB/OL ]. [ 2019-03-20 ]. <http://de.arxiv.org/pdf/1409.0473>.
- [ 6 ] SEE A, LIU P J, MANNING C D. Get to the point: summarization with pointer-generator networks [ C ] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Washington D. C., USA: IEEE Press, 2017: 1073-1083.
- [ 7 ] GEHRING J, AULI M, GRANGIER D, et al. A convolutional encoder model for neural machine translation [ EB/OL ]. [ 2019-03-20 ]. <http://wing.comp.nus.edu.sg/~antho/P/P17/P17-1012.pdf>.
- [ 8 ] LIN C Y, HOVY E. Automatic evaluation of summaries using N-gram co-occurrence statistics [ C ] // Proceedings of 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology. [ S. l. ] : ACL, 2003: 26-35.
- [ 9 ] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [ J ]. Neural Computation, 1997, 9 ( 8 ) : 1735-1780.
- [ 10 ] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [ EB/OL ]. [ 2019-03-20 ]. <http://de.arxiv.org/pdf/1406.1078>.
- [ 11 ] YOON K. Convolutional neural networks for sentence classification [ EB/OL ]. [ 2019-03-20 ]. <http://de.arxiv.org/pdf/1408.5882>.
- [ 12 ] MENG F, LU Z, WANG M, et al. Encoding source language with convolutional neural network for machine translation [ EB/OL ]. [ 2019-03-20 ]. <http://aclweb.org/anthology/P/P15/P15-1003.pdf>.
- [ 13 ] PANG Lei, LI Shoushan, ZHOU Guodong. Sentiment classification method of Chinese micro-blog based on emotional knowledge [ J ]. Computer Engineering, 2012, 38 ( 13 ) : 156-158, 162. ( in Chinese )  
庞磊,李寿山,周国栋.基于情绪知识的中文微博情感分类方法[ J ]. 计算机工程, 2012, 38 ( 13 ) : 156-158, 162.
- [ 14 ] CHO K, VAN MERRIËNBOER B, BAHDANAU D, et al. On the properties of neural machine translation: encoder-decoder approaches [ EB/OL ]. [ 2019-03-20 ]. <http://de.arxiv.org/pdf/1409.1259>.
- [ 15 ] YOUNG T, HAZARIKA D, PORIA S, et al. Recent trends in deep learning based natural language processing [ J ]. IEEE Computational Intelligence Magazine, 2018, 13 ( 3 ) : 55-75.
- [ 16 ] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks [ J ]. IEEE Transactions on Signal Processing, 1997, 45 ( 11 ) : 2673-2681.
- [ 17 ] ZHOU Feiyan, JIN Linpeng, DONG Jun. Review of convolutional neural network [ J ]. Chinese Journal of Computers, 2017, 40 ( 6 ) : 1229-1251. ( in Chinese )  
周飞燕,金林鹏,董军.卷积神经网络研究综述[ J ]. 计算机学报, 2017, 40 ( 6 ) : 1229-1251.
- [ 18 ] GOODFELLOW I J, WARDE-FARLEY D, MIRZA M, et al. Maxout networks [ J ]. Computer Science, 2013, 39 : 1319-1327.
- [ 19 ] ZHOU Q, YANG N, WEI F, et al. Neural question generation from text: a preliminary study [ C ] // Proceedings of National CCF Conference on Natural Language Processing and Chinese Computing. Berlin, Germany: Springer, 2017: 662-671.
- [ 20 ] GULCEHRE C, AHN S, NALLAPATI R, et al. Pointing the unknown words [ C ] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. [ S. l. ] : ACL, 2016: 256-263.
- [ 21 ] FREITAG M, AL-ONAIZAN Y. Beam search strategies for neural machine translation [ C ] // Proceedings of the 1st Workshop on Neural Machine Translation. [ S. l. ] : ACL, 2017: 56-60.
- [ 22 ] HU B, CHEN Q, ZHU F. LCSTS: a large scale Chinese short text summarization dataset [ J ]. Computer Science, 2015, 41 : 2667-2671.
- [ 23 ] KINGMA D P, BA J. Adam: a method for stochastic optimization [ EB/OL ]. [ 2019-03-20 ]. <https://arxiv.org/pdf/1412.6980.pdf>.

编辑 吴云芳