计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190600006

基于语义感知的中文短文本摘要生成模型

倪海清　刘　丹　史梦雨
电子科技大学电子科学技术研究院　成都 611731
(nihaijing0520@163.com)

摘　要　文本摘要生成技术能够从海量数据中概括出关键信息,有效解决用户信息过载的问题。目前序列到序列模型被广泛应用于英文文本摘要生成领域,而在中文文本摘要生成领域没有对该模型进行深入研究。对于传统的序列到序列模型,解码器通过注意力机制将编码器输出的每一个词的隐藏状态作为原始文本完整的语义信息来生成摘要,但是编码器输出的每一个词的隐藏状态仅包含前、后词的语义信息,不包含原始文本完整的语义信息,导致生成摘要缺失原始文本的核心信息,影响生成摘要的准确性和可读性。为此,文中提出基于语义感知的中文短文本摘要生成模型SA-Seq2Seq,以结合注意力机制的序列到序列模型为基础,通过使用预训练模型BERT,在编码器中将中文短文本作为整体语义信息引入,使得每一个词包含整体语义信息;在解码器中将参考摘要作为目标语义信息计算语义不一致损失,以确保生成摘要的语义完整性。采用中文短文本摘要数据集LCSTS进行实验,结果表明,模型SA-Seq2Seq在评估标准ROUGE上的效果相对于基准模型有显著提高,其ROUGE-1,ROUGE-2和ROUGE-L评分在基于字符处理的数据集上分别提升了3.4%,7.1%和6.1%,在基于词语处理的数据集上分别提升了2.7%,5.4%和11.7%,即模型SA-Seq2Seq能够更有效地融合中文短文本的整体语义信息,挖掘其关键信息,确保生成摘要的流畅性和连贯性,可以应用于中文短文本摘要生成任务。

关键词:中文短文本摘要;序列到序列模型;注意力机制;预训练模型;语义感知
中图法分类号　TP391.1

Chinese Short Text Summarization Generation Model Based on Semantic-aware

NI Hai-qing,LIU Dan and SHI Meng-yu
Research Institute of Electronic Science and Technology,University of Electronic Science and Technology of China,Chengdu 611731,China

Abstract　The text summary generation technology can summarize the key information from the massive data and effectively solve the problem of information overload. At present,the sequence-to-sequence model is widely used in the field of English text abstraction generation,but there is no in-depth study on this model in the field of Chinese text abstraction. In the conventional sequence-to-sequence model,the decoder applies the hidden state of each word output by the encoder as the overall semantic information through the attention mechanism,nevertheless the hidden state of each word which encoder outputs only in consideration of the front and back words of current word,which results in the generated summary missing the core information of the source text. To solve this problem,a semantic-aware based Chinese short text summarization generation model called SA-Seq2Seq is proposed,which uses the sequence-to-sequence model with attention mechanism. The model SA-Seq2Seq applies the pre-training model called BERT to introduce source text in the encoder so that each word contains the overall semantic information and uses gold summary as the target semantic information in the decoder to calculate the semantic inconsistency loss,thus ensuring the semantic integrity of the generated summary. Experiments are carried out on the dataset using the Chinese short text summary dataset LCSTS. The experimental results show that the model SA-Seq2Seq on the evaluation metric ROUGE is significantly improved compared to the benchmark model,and its ROUGE-1,ROUGE-2 and ROUGE-L scores increase by 3.4%,7.1% and 6.1% respectively in the dataset that is processed based on character and increase by 2.7%,5.4% and 11.7% respectively in the dataset that is processed based on word. So the SA-Seq2Seq model can effectively integrate Chinese short text and ensure the fluency and consistency of the generated summary,which can be applied to the Chinese short text summary generation task.

Keywords　Chinese short text summarization,Sequence to sequence model,Attention mechanism,Pre-training model,Semantic aware

到稿日期:2019-06-01　返修日期:2019-08-11　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
通信作者:刘丹(liudan@uestc.edu.cn)

1 引言

随着信息化时代的到来，人们无时无刻地创造信息，造成了信息过载的问题。如何从海量的信息中获取有价值的信息成为一个亟待解决的问题。文本摘要被认为是解决该问题的关键技术，它能够有效地概括出文本中的重要信息。

文本摘要按照实现方式的不同分为抽取式文本摘要和生成式文本摘要。抽取式文本摘要[1]是从原始文本中抽取一些关键词、中心句后组合成摘要；而生成式文本摘要[2]是根据原始文本的核心思想，通过句子压缩、同义替换等方式生成对应的摘要。

随着深度神经网络模型的发展，序列到序列模型[3](Sequence to Sequence)在文本摘要生成任务中得到了广泛的应用。文本摘要生成的关键问题是生成摘要存在词语重复、语义损失等问题。Bahdanau 等[4]在序列到序列模型的基础上提出了注意力机制，在不同的时刻输入不同的语义向量。Luong 等[5]在序列到序列模型的基础上通过全局注意力机制引入所有的词，同时使用局部注意力机制减少过长的输入文本的计算量。Pang 等[6]在序列到序列模型上引入分类器，充分利用监督信息来获得更多的摘要特性。Wu 等[7]在序列到序列模型上，设计全局匹配机制，仅在编码器端引入全局信息来生成摘要。而现阶段随着预训练模型的兴起，例如 Word2Vector[8]，ELMO[9]，GPT[10](或 GPT2[11])，BERT[12]，有些研究工作将预先训练的模型引入文本摘要生成任务，Kryściński 等[13]基于参考摘要预先训练语言模型，在解码器中使用预训练的语言模型作为先验知识。

当前主流的序列到序列模型不能有效地对长文本进行语义编码，相关研究也集中在短文本的摘要生成。除此之外，由于中文和英文的语法结构和表现形式存在着较大差异，使得对中文文本摘要生成问题的研究比较复杂。对于传统的序列到序列模型，解码器通过注意力机制将编码器输出的每一个词的隐藏状态作为原始文本完整的语义信息来生成摘要，但是编码器输出的每一个词的隐藏状态仅包含前、后词的语义信息，不包含原始文本的完整语义信息，导致生成摘要缺失原始文本的核心信息。

针对上述问题，本文研究中文短文本的摘要生成技术，提出了基于语义感知的中文短文本摘要生成模型 SA-Seq2Seq，以基于注意力机制的序列到序列模型为基础，通过使用预训练模型 BERT(Bidirectional Encoder Representations from Transformers)，在编码器中将中文短文本作为整体语义信息引入，同时在解码器中将参考摘要作为目标语义信息计算语义不一致损失，从而确保生成的摘要的语言流畅性和语义完整性。

2 基于语义感知的中文短文本摘要生成模型

如图 1 所示，模型 SA-Seq2Seq 的整体结构主要包括基于门控循环单元[14]的双向循环神经网络、基于注意力的编码器-解码器结构和语义感知层。其中，语义感知层的作用如下：1)将中文短文本作为整体语义信息引入编码器输出的每一个词的隐藏状态中；2)将参考摘要作为目标语义信息引入解码器，用于计算语义不一致损失。

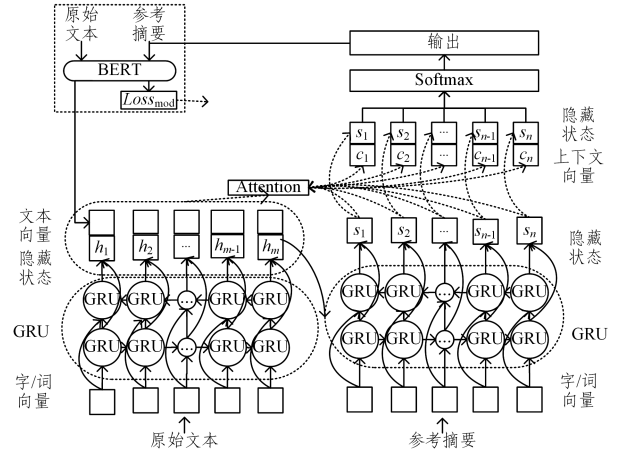


图 1 模型 SA-Seq2Seq 的结构

Fig. 1 Structure of SA-Seq2Seq model

SA-Seq2Seq 生成摘要的过程如下：编码器依次读取中文短文本中的每一个词，引入中文短文本的语义信息输出每一个词的隐藏状态；然后解码器通过引入参考摘要的语义信息后计算语义不一致损失来生成摘要，以确保生成摘要的语义完整性。

2.1 问题定义

在模型 SA-Seq2Seq 中，定义中文短文本 $X=(x_1,\cdots,x_m)$，生成摘要 $Y=(y_1,\cdots,y_n)$，参考摘要 $Y^*=(y_1^*,\cdots,y_m^*)$，则文本摘要生成任务的目标函数为：

$$p(X|Y) \tag{1}$$

2.2 双向循环神经网络

模型 SA-Seq2Seq 的编码器和解码器是基于门控循环单元的双向循环神经网络，它的基本思想是每一个训练序列向前和向后分别是两个标准的循环神经网络，使得输出的每一个词的隐藏状态提供包含完整的前、后词语的语义信息，其公式如下：

$$\overrightarrow{h_t}=f(x_t,\overrightarrow{h_{t-1}}) \tag{2}$$

$$\overleftarrow{h_t}=f(x_t,\overleftarrow{h_{t-1}}) \tag{3}$$

$$h_t=concat(\overrightarrow{h_t},\overleftarrow{h_t}) \tag{4}$$

其中，$\overrightarrow{h_t}$ 表示时刻 t 向前循环神经网络输出的隐藏状态；$\overleftarrow{h_t}$ 表示时刻 t 向后循环神经网络输出的隐藏状态；h_t 表示时刻 t 双向循环神经网络输出的隐藏状态；$concat$ 表示拼接每一个词的向前、向后的隐藏状态；f 表示非线性转换函数，可以是长短期记忆网络[14]或者门控循环单元[15]等。而在模型 SA-Seq2Seq 中使用的门控循环单元的公式如下：

$$z_t=\sigma(W_z[h_{t-1},x_t]) \tag{5}$$

$$r_t=\sigma(W_r[h_{t-1},x_t]) \tag{6}$$

$$\tilde{h}_t=\tanh(W[r_t h_{t-1},x_t]) \tag{7}$$

$$h_t=(1-z_t)h_{t-1}+z_t\tilde{h}_t \tag{8}$$

其中，z_t 表示更新门，r_t 表示重置门，h_t 表示 t 时刻门控循环单元输出的隐藏状态。

2.3 基于注意力机制的编码器-解码器结构

模型 SA-Seq2Seq 是基于注意力的编码器-解码器结构。

编码器从中文短文本 X 依次读取每一个词 $x_i(i=1,\cdots,m)$，生成对应的隐藏状态 $h_i(i=1,\cdots,m)$。

结合注意力机制[5]，解码器生成摘要 Y 中的每一个词 $y_j(j=1,\cdots,n)$都有对应的上下文向量 $\boldsymbol{c}_j(j=1,\cdots,n)$，该上下文向量是对编码器中的所有隐藏状态 $h_i(i=1,\cdots,m)$的加权，用于引入中文短文本中每一个词的语义信息，其公式如下：

$$e_{ij}=s_i^{\mathrm{T}}Wh_j \tag{9}$$

$$\alpha_{ij}=\frac{\exp(e_{ij})}{\sum_{i=1}^{m}\exp(e_{ij})} \tag{10}$$

$$\boldsymbol{c}_j=\sum_{i=1}^{m}\alpha_{ij}h_i \tag{11}$$

其中，e_{ij} 表示与每一个隐藏状态 $h_j(j=1,\cdots,n)$的权重系数；W 表示可训练的参数；e_{ij} 表示每一个词 $y_j(j=1,\cdots,n)$对每一个隐藏状态 $h_i(i=1,\cdots,m)$归一化后的权重系数。

最后，解码器拼接每一个词对应的隐藏状态 s_j 和上下文向量 $\boldsymbol{c}_j(j=1,\cdots,n)$，使得解码器具有完整的上下文信息，通过归一化指数函数来预测每一个词 $y_j(j=1,\cdots,n)$的生成概率，其公式如下：

$$y_j=softmax(MLP(concat(s_j,\boldsymbol{c}_j))) \tag{12}$$

其中，$softmax$ 表示归一化指数函数，MLP 表示全连接层，$concat$ 表示拼接隐藏状态 s_j 和上下文向量 $\boldsymbol{c}_j$。

2.4 语义感知层

模型 SA-Seq2Seq 的语义感知层的功能是动态地引入中文短文本和参考摘要的语义信息，分为编码和解码两个阶段：1)在编码阶段，通过预训练模型 BERT 将中文短文本作为整体语义信息引入；2)在解码阶段，通过预训练模型 BERT 将参考摘要作为目标语义信息用于计算语义不一致损失。

2.4.1 预训练模型

预训练模型在下游任务中的应用方式有两种：1)基于特征的方式，即将预训练模型的表示作为额外的特征引入下游任务；2)基于微调的方式，即根据特定的下游任务对预训练模型的参数进行小幅度的调整。

预训练模型 BERT[12]采用了 Transformer[16]编码器的模型作为语言模型，第一个阶段是语言模型的预训练，共有两个预训练任务：1)完形填空(Masked Language Model)，它基于词语序列中其他没有遮蔽的单词提供的上下文，预测被遮蔽的词语；2)下一句预测(Next Sentence Prediction)，在句子层面充分挖掘句子层面的语义信息，即预测第二句话是否是第一句话的下一句话。其第二个阶段是基于微调的方式解决下游任务。

本文提出的模型 SA-Seq2Seq 基于特征的方式使用预训练模型 BERT，即通过预训练模型 BERT 对中文短文本和参考摘要进行向量化表示，分别作为整体语义信息和目标语义信息。

2.4.2 整体语义信息

模型 SA-Seq2Seq 的编码器依次读取中文短文本 X 中的每一个词 $x_i(i=1,\cdots,m)$，生成的隐藏状态 $h_i(i=1,\cdots,m)$仅仅包含了当前词的前、后词的语义信息，因此包含中文短文本的完整语义信息，那么通过预训练模型 BERT 将中文短文本作为整体语义信息引入。

首先将中文短文本 $X=(x_1,\cdots,x_m)$转换为以词为单位的序列 $W_X=(w_1,\cdots,w_m)$，输入预训练模型 BERT 进行文本向量化后，得到整体语义信息 d_{text}，然后修改式(4)，将其引入输出的每一个词的隐藏状态，其对应的公式如下：

$$d_{\text{text}}=Bert(W) \tag{13}$$

$$h_{\text{text}}=MLP(d_{\text{text}}) \tag{14}$$

$$h_t=concat(\overrightarrow{h_t},\overleftarrow{h_t},h_{\text{text}}) \tag{15}$$

其中，$Bert$ 表示通过预训练模型 BERT 得到的整体语义信息；MLP 表示多层感知机，用于调整整体语义信息与编码器输出的隐藏状态维度一致；$concat$ 表示拼接向前、向后隐藏状态和整体语义信息。

2.4.3 目标语义信息

模型 SA-Seq2Seq 的解码器中，根据每一个词的隐藏状态 $s_j(j=1,\cdots,n)$和上下文向量 $\boldsymbol{c}_j(j=1,\cdots,n)$生成摘要 Y，然后计算损失值 $Loss_{\text{dec}}$，其公式如下：

$$Loss_{\text{dec}}=-\log(\sum_{j=1}^{n}p(y_j\mid s_j,c_j)) \tag{16}$$

为了使生成摘要 Y 与参考摘要 Y^* 在语义上对齐，引入语义不一致损失。

首先分别将生成摘要 Y 和参考摘要 Y^* 转换为以词为单位的序列 $W_Y=(w_1,\cdots,w_n)$和 $W_{Y^*}=(w_1^*,\cdots,w_m^*)$，然后通过预训练模型 BERT 获得对应语义向量 $\boldsymbol{d}_{\text{gen}}$和目标语义信息 $\boldsymbol{d}_{\text{ref}}$，最后使用余弦相似性计算语义不一致损失 $Loss_{\text{sim}}$，其计算公式如下：

$$\boldsymbol{d}_{\text{gen}}=Bert(Y) \tag{17}$$

$$\boldsymbol{d}_{\text{ref}}=Bert(Y^*) \tag{18}$$

$$Loss_{\text{sim}}=\frac{\boldsymbol{d}_{\text{gen}}\cdot\boldsymbol{d}_{\text{ref}}}{\|\boldsymbol{d}_{\text{gen}}\|\ \|\boldsymbol{d}_{\text{ref}}\|} \tag{19}$$

其中，$Bert$ 表示通过预训练模型 BERT 得到语义信息，$\|\cdot\|$ 表示向量的模。

2.4.4 训练和预测

在模型 SA-Seq2Seq 训练时，在解码器中为减少误差的累积，使用导师驱动(Tearch Forcing)算法[17]依次输入参考摘要中每一个词 Y^* 中的每一个词 y_j^*，生成对应的隐藏状态 $s_j(j=1,\cdots,n)$，用于计算损失值 $Loss_{\text{dec}}$。而在引入语义不一致损失值 $Loss_{\text{sim}}$后，模型 SA-Seq2Seq 的损失函数的公式如下：

$$Loss_{\text{mod}}=Loss_{\text{dec}}+Loss_{\text{sim}} \tag{20}$$

在模型 SA-Seq2Seq 测试时，在解码器中使用集束搜索(Beam-Search)算法，它是一种启发式图搜索算法，该算法使用广度优先策略建立搜索树，在树的每一层按照一定的评价代价对节点进行排序，保留预先设定集束宽度(Beam Width)的节点，仅在这些节点上进行下一层的拓展，即在生成摘要输出前检索大量的词汇，从而得到最优的选择。

3 实验

3.1 实验数据集

本文实验使用大规模中文短文摘要数据集(LCSTS)[18]，该数据来源于中国社交网络新浪微博，共分为 3 个部分，第一部分共有 240 万条数据，第二部分共有 1 万多条数据，第三部分共有 0.1 万多条数据，其中每条数据包含一个中文短文本和一个相对应的摘要，第二部分和第三部分还包含中文短文

本与对应摘要的相关性。评分范围为1～5，表示中文短文本和对应摘要之间的相关性，“1”表示“最不相关”，“5”表示“最相关”。为了更好地加强模型的泛化能力，将第一部分数据作为训练数据集，将第三部分中相关性评分为3，4，5的数据作为测试集。对实验数据进行预处理，具体如下：

(1)去除文本中的特殊字符和标点符号；

(2)使用标签NUM替换所有的数字，使用标签DATA替换所有的日期；

(3)基于字符、词语对文本进行预处理，其中基于词语的处理使用结巴分词。

3.2 评估标准

ROUGE评估标准[19]是一种广泛使用的文本摘要评估方法，一般使用ROUGE-N和ROUGE-L进行文本摘要评价。本文使用ROUGE-1、ROUGE-2和ROUGE-N评价模型SA-Seq2Seq生成摘要的效果。

ROUGE-N的计算式如下：

$$ROUGE\text{-}N=\frac{\sum_{gram_N\in Y}Count_M(gram_N)}{\sum_{gram_N\in Y}Count_Y(gram_N)} \tag{21}$$

其中，$gram_N$表示N-$gram$；$Count_Y(gram_N)$表示Y中的最大共现数量N-$gram$，$Count_M(gram_N)$表示N-$gram$在生成摘要中的最大共现数量。

ROUGE-L的计算式如下：

$$ROUGE\text{-}L=\frac{(1+\beta^2)r_Lp_L}{r_L+\beta^2p_L} \tag{22}$$

$$r_L=\frac{LCS(Y^*,Y)}{Len(Y)} \tag{23}$$

$$p_L=\frac{LCS(Y^*,Y)}{Len(Y^*)} \tag{24}$$

其中，$LCS(Y^*,Y)$表示生成摘要Y^*和参考摘要Y之间的最长公共子序列，r_L表示召回率，p_L表示准确率。

3.3 实验参数

为了更好地验证模型的效果，基于字符处理的字典大小和基于词语的词典大小与Hu等[18]的实验中的字典、词典大小一致。根据统计字符和词语出现的频次，基于字符处理的字典大小为4 000，基于词语处理的词典大小为50 000，同时使用UNK表示未出现字典中的字符或者词典中的词语；模型使用Adam优化器[20]；编码器和解码器的隐藏层节点的个数为256；预训练模型BERT[10]使用预先训练好的模型BERT-Base(Chinese)；在SA-Seq2Seq模型进行解码时，集束搜索算法中的集束宽度设置为10。

3.4 模型说明

为了评估提出的模型SA-Seq2Seq在中文短文本摘要任务中的效果，引入Hu等[18]提出的模型RNN与模型RNN-Con作为参照。为了更好地评估模型SA-Seq2Seq的效果，本文单独实现了基于注意力机制的序列到序列的模型Base-Seq2Seq，将其作为基准模型与模型SA-Seq2Seq进行比较。

实验模型的说明如下：

(1)RNN[18]：循环神经网络作为编码器，将其最后的隐藏状态输入解码器；

(2)RNN-Con[18]：循环神经网络作为编码器，将所有隐藏状态的组合输入解码器；

(3)Base-Seq2Seq：基于注意力机制的序列到序列模型，没有引入整体语义信息和目标语义信息；

(4)SA-Seq2Seq：基于注意力机制的序列到序列模型，同时引入整体语义信息和目标语义信息。

3.5 实验结果及分析

表1中每一个单元中的数据分别是对数据集基于字符和词语处理后，对应模型生成摘要评估获得的值。通过分析可知，基准模型Base-Seq2Seq，即单独实现的模型的效果优于Hu等[18]提出的模型RNN与模型RNN-Con，其ROUGE-1，ROUGE-2和ROUGE-L的值在基于字符处理的数据集上分别提升了9.4%，21.3%和8.5%，在基于词语处理的数据集上分别提升了23.9%，36.0%和24.1%，其中基于词语处理的效果优于基于字符处理的效果。在引入整体语义信息和目标语义信息后，模型SA-Seq2Seq的效果优于基准模型，其ROUGE-1，ROUGE-2和ROUGE-L的值在基于字符处理的数据集上分别提升了3.4%，7.1%和6.1%，在基于词语处理的数据集上分别提升了2.7%，5.4%和11.7%。

表1 每个模型在数据集上的实验结果

Table 1 Experiment results of each model on dataset

模型	ROUGE-1	ROUGE-2	ROUGE-L
RNN[18]	17.7/21.5	8.5/8.9	15.8/18.6
RNN-Con[18]	29.9/26.8	17.4/16.1	27.2/24.1
Base-Seq2Seq	32.7/33.2	21.1/21.9	29.5/29.9
SA-Seq2Seq	33.8/34.1	22.6/23.1	31.3/33.4

如表2所列，Base-Seq2Seq-c和SA-Seq2Seq-c是在基于字符处理的数据集上生成的摘要；Base-Seq2Seq-w和SA-Seq2Seq-w是在基于词语处理的数据集上生成的摘要。通过对比分析可得，模型SA-Seq2Seq生成的摘要效果较好，对中文短文本提取出包含较多语义信息的文本摘要，虽然其生成的摘要存在着字符或词语重复，但是在语义上具有一定流畅性。另外，在基于词语处理的词典上，SA-Seq2Seq生成的摘要效果较好。

表2 生成摘要的样例

Table 2 Examples of generation summarization

原始文本	央行今日将召集大型商业银行和股份制银行开会，以应对当前的债市风暴。消息人士表示，央行一方面旨在维稳银行间债券市场，另一方面很可能探讨以两类户治理为重点的改革内容。此次债市风暴中，国家审计署扮演了至关重要的角色
参考摘要	媒体称央行今日召集银行开会应对当前债市风暴
Base-Seq2Seq-c	媒体体称集级银银开会应对当前的市场的风暴
SA-Seq2Seq-c	媒体体称行召集银行开应对当前债市市场风暴
Base-Seq2Seq-w	媒体媒体称央行央行银行开会应对当前的的市场风暴
SA-Seq2Seq-w	媒体央行召集银行开会应对应对当前债市场市场风暴

结束语 在传统的序列到序列模型中，解码器通过注意力机制将编码器输出的每一个词的隐藏状态作为整体语义信息引入，但是编码器输出的每一个词的隐藏状态仅考虑前、后词的语义信息，导致生成摘要缺失原始文本的核心信息。针对上述问题，本文提出了一个基于语义感知的中文短文本摘要生成模型SA-Seq2Seq，以基于注意力机制的序列到序列模型为基础，引入整体语义信息和目标语义信息，使得生成的摘要具有语言流畅性和语义完整性。

实验表明，模型SA-Seq2Seq能够更有效地融合中文短文

本的整体语义信息，挖掘出其关键信息。后续工作将在不同的数据集上进行测试，检验模型 SA-Seq2Seq 的有效性。除此之外，虽然预训练模型 BERT 具有强大的语义表示能力，但是直接使用预训练模型 BERT 对中文短文本和参考摘要进行向量化表示，仍存在一定的语义损失，需要进一步研究如何基于预训练模型 BERT 充分地对中文短文本和参考摘要进行向量化表示，使其包含更为详实的语义信息，改善模型 SA-Seq2Seq 生成文本摘要的效果。

参 考 文 献

[1] NETO J L, FREITAS A A, KAESTNER C A A. Automatic Text Summarization Using a Machine Learning Approach[C]// 16th Brazilian Symposium on Artificial Intelligence. 2002: 11-14.

[2] CHOPRA S, AULI M, RUSH A M. Abstractive Sentence Summarization with Attentive Recurrent Neural Networks[C]// Conference of the North American Chapter of the Association for Computational Linguistics. 2016: 93-98.

[3] SUTSKEVER I, VINYALS O, LE Q V. Sequence to Sequence Learning with Neural Networks[C]// Advances in Neural Information Processing Systems 27. 2014: 3104-3112.

[4] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[C]// 3rd International Conference on Learning Representations. 2015.

[5] LUONG M T, PHAM H, MANNING C D. Effective Approaches to Attention-based Neural Machine Translation[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 1412-1421.

[6] PANG C, YIN C H. Chinese Text Summarization Based on Classification[J]. Computer Science, 2018, 45(1): 144-147, 178.

[7] WU R S, WANG H L, WANG Z Q, et al. Short Text Summary Generation with Global Self-Matching Mechanism[J/OL]. Journal of Software. https://doi.org/10.13328/j.cnki.jos.005850.

[8] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality[C]// 27th Annual Conference on Neural Information Processing Systems 2013. 2013: 3111-3119.

[9] PETERS M E, NEUMANN M, IYYER M, et al. Deep Contextuaized Word Representations[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018: 2227-2237.

[10] RADFORD A, NARASIMHAN K, et al. Improving language understanding by generative pre-training[EB/OL]. https://s3-us-west-2.amazonaws.com/openaiassets/research-covers/language-unsupervised/languageunderstandingpaper.pdf.

[11] RADFORD A, WU J, et al. Language Models are Unsupervised Multitask Learners[EB/OL]. https://blog.openai.com/better-language-models/.

[12] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. CoRR, 2018, abs/1810.04805.

[13] KRYŚCINSKI W, PAULUS R, XIONG C, et al. Improving Abstraction in Text Summarization[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 1808-1817.

[14] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[15] CHO K, VAN MERRIENBOER B, BAHDANAU D, et al. On the Properties of Neural Machine Translation: Encoder--Decoder Approaches[C]// Proceedings of SSST@EMNLP 2014. 2014: 103-111.

[16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All you Need[C]// Advances in Neural Information Processing Systems. 2017: 6000-6010.

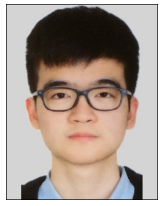
[17] WILLIAMS R J, ZIPSER D. A learning algorithm for continually running fully recurrent neural networks[J]. Neural Computation, 1989, 1(2): 270-280.

[18] HU B, CHEN Q, ZHU F, et al. LCSTS: A Large Scale Chinese Short Text Summarization Dataset[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015: 1967-1972.

[19] LIN C. ROUGE: A Package for Automatic Evaluation of Summaries[C]// Text Summarization Branches Out: Proceedings of the ACL-04 Workshop. 2004: 74-81.

[20] KINGMA D P, BA J. Adam: A method for stochastic optimization[C]// 3rd International Conference on Learning Representations. 2014.

[21] NETO J L, FREITAS A A, KAESTNER C A, et al. Automatic Text Summarization Using a Machine Learning Approach[C]// Brazilian Symposium on Artificial Intelligence. 2002: 205-215.



NI Hai-qing, born in 1994, postgraduate. His main research interests include text summarization and so on.

LIU Dan, born in 1969, Ph.D, associate professor. His main research interests include network and system security, cloud computing and data processing.