计算机应用研究
Application Research of Computers
ISSN 1001-3695,CN 51-1196/TP

《计算机应用研究》网络首发论文

题目：融合复制机制和 input-feeding 方法的中文自动摘要模型
作者：农丁安，欧阳纯萍，阳小华
DOI：10.19734/j.issn.1001-3695.2019.03.0065
收稿日期：2019-03-01
网络首发日期：2019-07-24
引用格式：农丁安，欧阳纯萍，阳小华. 融合复制机制和 input-feeding 方法的中文自动摘要模型[J/OL]. 计算机应用研究.
https://doi.org/10.19734/j.issn.1001-3695.2019.03.0065

网络首发时间：2019-07-24 15:22:55
网络首发地址：http://kns.cnki.net/kcms/detail/51.1196.tp.20190723.1730.005.html

融合复制机制和 input-feeding 方法的中文自动摘要模型 *

农丁安，欧阳纯萍†，阳小华

(南华大学 计算机学院，湖南 衡阳 421001)

摘 要：针对中文自动摘要准确率不高的问题，在含有注意力机制的序列到序列(seq2seq)基础模型的解码器中融合了复制机制和 input-feeding 方法，提出了准确率更高的中文自动摘要模型。首先，该模型使用指针网络将出现在源序列中的 OOV(out-of-vocabulary)词扩展到固定词典，以实现从源序列复制 OOV 词到生成序列中；其次，input-feeding 方法用于跟踪已生成序列的注意力决定信息以提升模型输出准确率。在 NLPCC2018 数据集上的实验结果表明，与基础模型相比，所提出模型获得了更高的 ROUGE 得分，验证了该方法的可行性。

关键词：自动摘要；复制机制；input-feeding 方法；指针网络；序列到序列；注意力机制
中图分类号：TP391.1 doi: 10.19734/j.issn.1001-3695.2019.03.0065

Chinese automatic summarization model of combining copying mechanism and input-feeding approach

Nong Ding'an, Ouyang Chunping†, Yang Xiaohua

(*School of Computer, University of South China, Hengyang Hunan* 421001, *China*)

Abstract: This paper presented a novel model for the lower accuracy issue in Chinese automatic summary which merged copying mechanism and input-feeding approach into the decoder of sequence-to-sequence (seq2seq) basic model with attention mechanism. Firstly, it used Pointer Networks to extend the source's out-of-vocabulary (OOV) words to a fixed dictionary to copy OOV words from the source into the generated sequence. Secondly, it used the input-feeding approach to track the attention decision information of generated sequence for improving the model output accuracy. Experimental results on NLPCC2018 datasets show that the proposed model obtains a higher ROUGE score than the basic model, which confirms the feasibility of this method.

Key words: automatic summarization; copying mechanism; input-feeding approach; pointer networks; sequence-to-sequence(seq2seq); attention mechanism

0 引言

随着互联网技术的不断发展，数据量也在不断增大，信息过载现象开始变得明显。如何有效解决人们快速、准确地获取文本中的主题信息已成为自然语言处理领域的一个研究热点。自动摘要作为一种文本解释的重要工具，其通过对源文本进行压缩和精炼，可提取出能概括源文本主题的关键信息，为下游应用(如新闻摘要、搜索引擎以及报告生成等领域)提供有效支持，极大提高了用户获取信息的效率。通常，自动摘要可分为抽取式摘要和生成式摘要两类[1]，抽取式摘要指的是不对源文本句子进行修改，而直接抽取出关键句子组成摘要，基本思想是先通过一定的方法对源文本中每个句子计算其得分，选出得分较高的句子组成摘要；而生成式摘要则是计算机通过对源文本进行充分理解后，对文本进行抽象，然后使用语义相近的词或短语重新组织句子生成摘要，其形式类似于人工编写的摘要，更符合人们的阅读习惯。自动摘要现已成为自然语言处理领域的研究热点之一。

1 相关工作

自动摘要传统的研究技术是基于抽取式方法。Carbonell 等人将相关性和新颖性相结合，提出了最大边际相关(maximal marginal relevance，MMR)算法[2]，该方法可有效减少摘要的冗余；Mihalcea 等人提出了一种图排序的 TextRank 算法[2]，与其他抽取式方法相比，所构建的模型优势更明显；Erkan 等人使用图对句子进行表示[2]，通过特征向量中心性的概念计算句子重要性，在 DUC 2004 数据集上评估得到了排名第一的成绩；余珊珊等人[3]将标题、段落等信息引入 TextRank 网络图中，提出的 iTextRank 算法提高了抽取式方法的摘要准确率；刘彼洋等人[4]提出了一种将矩阵分解与子模最大化相结合的面向微博短文本自动摘要方法，有效改善了基线系统的摘要性能。

近年来，基于神经网络的编码器-解码器模型在解决 seq2seq[5]问题上表现优异，如机器翻译[6,7]、文本生成[8]等任务都在这类模型上取得了成功。因此，基于神经网络的生成式方法已成为现今自动摘要的研究热点，Rush 等人[9]将词袋模型(bag-of-words)、CNN(convolutional neural networks，CNN)和基于注意力(attention)的三种不同编码器用于编码源序列，在 DUC 2004 数据集上得到了基准摘要模型；Chopra 等人则采用基于注意力的卷积编码器[10]，实现了解码过程中目标序列和源序列词的对齐，改善了文献[9]的基准模型；Hu 等人构建了一个大型中文短文本自动摘要数据集(LCSTS)[11]，并为该数据集设计了模型作为评价基准；Li 等人则提出一种具有递归生成解码的模型[12]，以学习目标摘要中隐含的潜在结构信息，有效改善了 LCSTS 中的基准模型。最近，Gu 等人

收稿日期：2019-03-01；**修回日期：**2019-04-16 **基金项目：**国家自然科学基金(61402220，61502221)；湖南省哲学社会科学基金(16YBA323)；湖南省自然科学基金项目(2015JJ3015)；湖南省教育厅青年项目(15B207)

作者简介：农丁安(1992-)，男，广西钦州人，硕士研究生，主要研究方向为信息检索与自然语言处理；欧阳纯萍(1979-)，女(通信作者)，湖南衡阳人，教授，博士，主要研究方向为语义网与社交网络(ouyangcp@126.com)；阳小华(1963-)，男，湖南衡阳人，博导，教授，主要研究方向为信息检索与知识发现.

[13]基于指针网络(pointer networks)提出了 CopyNet 模型[14]，首次将复制机制融合到 RNn(recurrent neural networks，RNN)的 seq2seq 学习中，通过复制源序列中的词或短语到生成序列中，有效实现了抽取式和生成式两种方法相结合，将 LCSTS 的基准模型最多提高了约 8%；应文豪等人[15]通过分析句子与话题的语义关系并构建了基于排序学习的半监督训练框架，在 DUC2004 多文档摘要任务上有效改进了摘要质量。

基于神经网络的生成式方法在生成摘要时，需要先构建一个固定大小的词典，从而根据词典中每个词的概率大小来选择最终词输出。然而，如果某个关键词不在词典中，那么它在词典中的概率为零，网络将会忽略这个词，导致许多不在词典中的关键词无法生成，这会大大降低模型的准确率；同时，在每个解码时刻，虽然使用注意力机制(attention mechanism)可以对编码器中的每个隐层状态进行动态关注，但是每个时刻的注意力都是相互独立的，各个时刻只独立根据各自的注意力信息作出当前时刻的决定，而没有参考历史时刻的注意力决定信息，这会降低注意力决定的准确率。

针对上述问题，提出了将复制机制和 input-feeding 方法融入到 seq2seq 架构中来提升模型输出摘要的准确率。首先，复制机制借鉴了 See 等人[16]对指针网络的改进，用于从源序列中复制词或短语作为目标输出；其次，input-feeding 方法依赖于文献[7]的 GlobalAttention 生成的注意力向量，作用是将前一时刻的注意力决定传送到当前时刻，使得模型在每个时刻的注意力决定更加准确。通过将以上两个方法融合得到的模型在 NLPCC2018 数据集上进行的实验发现，与抽取式方法的典型模型相比，ROUGE 得分得到了大幅提高；同时，与文献[16]的指针网络模型相比，分别将 char 级别模型和 word 级别模型的 ROUGE 得分最多提高了 2.22%和 2.96%。说明使用这种融合的方法可有效提升中文自动摘要准确率。

2　中文自动摘要模型

模型的主体由编码器、解码器、注意力层和复制软转换(soft switch)四部分组成，结构如图 1 所示。其中，编码器为一个单层的 Bi-LSTM[17](Bi-directional long short-term memory, Bi-LSTM)用于编码源序列；解码器为一个单层 LSTM 用于解码输出目标序列；注意力层的作用是对编码隐层状态进行全局动态关注；复制软转换用于决定是从源序列中复制一个词还是由网络自主生成一个词，目的用于解决 OOV 词问题。模型训练结束后，使用文献[5]的 beam-search 算法进行解码推断输出摘要。

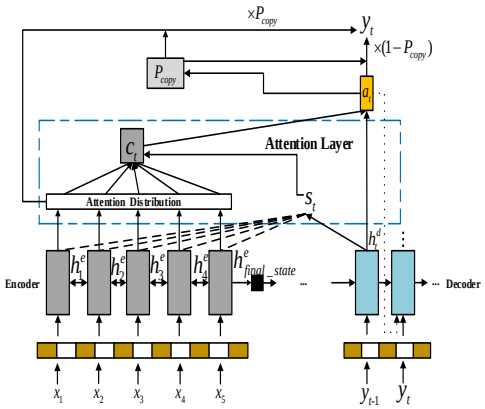

图 1　中文自动摘要模型

Fig. 1　Chinese automatic summarization model

为了能更好说明本文模型的各个方法在自动摘要中的实际应用。引用了如图 2 所示的实例进行说明。

Article:京华 时报： 天津 将 于 5 月底 取消 蓝印 户口 政策 ， “ 积分 落户 ” 将 取而代之 ， 这 意味着 已有 20 年 历史 的 蓝印 户口 将 退出 天津 历史舞台 .....
Summarization:天津 ： 5 月 31 日起 外地人 买房 不再 送 户口 ， 取消 蓝印 户口 政策 ， 以 “ 积分 落户 ” 取代 。

图 2　方法实例说明

Fig. 2　Method example description

如图 2 所示，使用注意力机制后，解码器在进行解码时，对源序列中的某些词将会重点关注(颜色越深)，而对某些词的关注程度较少(颜色越浅)。同时，例如“蓝印”这个词没有被包含在词典中，但它出现在源序列中，则使用复制机制可将该词复制到生成序列，从而实现了 OOV 词的生成。而 input-feeding 方法类似人类在编写摘要时，需要时刻考虑已编写的部分摘要，最终才能写出准确而完整的摘要。

2.1　编码器

设编码器共有 N 个编码时刻($X=\{x_1,x_2,...,x_N\}$)，第 i 个输入字符为 x_i，对应的词嵌入向量为 $embx_i$，则 x_i 在编码器中的前向和后向隐层状态分别为

$$h_t^e(fw)=BiLSTM(embx_t,h_{t-1}^e(fw)) \tag{1}$$

$$h_t^e(bw)=BiLSTM(embx_t,h_{t-1}^e(bw)) \tag{2}$$

h_{t-1}^e 是上一时刻编码隐层状态，将前向状态和后向状态拼接起来最终得到的 h_t^e 为

$$h_t^e=cat[h_t^e(fw),h_t^e(bw)] \tag{3}$$

其中，$cat[\bullet,\bullet]$ 表示将两个向量拼接。编码器在每个时刻都会生成一个隐层状态，这些状态还将用于解码阶段计算上下文向量；同时将最后一个时刻的编码状态 $h_{final\_state}^e$ 作为解码器的初始化状态。从而完成了对源序列的编码表示。

2.2　解码器

在解码器中使用加入了 GlobalAttention 层，它在计算每个解码时刻所对应的注意力分布时，都会全局考虑编码器中每个词的隐层状态。具体的，假设 t 时刻 LSTM 单元输出的解码隐层状态为 h_t^d，则 h_t^d 的求解为

$$h_t^d=LSTM(h_{t-1}^d,emby_t) \tag{4}$$

h_{t-1}^d 表示 $t-1$ 时刻的解码隐层状态，$emby_t$ 为 t 时刻 y_t 输入的词嵌入向量。接着使用一个 $score$ 函数将 t 时刻的解码状态 h_t^d 与第 i 个输入位置的编码状态 h_i^e 实现对齐：

$$e_{t,i}=score(h_t^d,h_i^e) \tag{5}$$

这里选择了 $h_t^dW_eh_i^e$ 作为 $score$ 函数，$e_{t,i}$ 是对齐向量，对应了两个向量的相似度大小。对 $e_{t,i}$ 进行 $soft\max$ 归一化处理，得到解码时刻 t 对 h_i^e 的注意力大小为

$$s_{t,i}=\frac{\exp(e_{t,i})}{\sum_{k=1}^{N}\exp(e_{t,k})} \tag{6}$$

根据式(6)求得解码时刻 t 的上下文向量 c_t 为

$$c_t=\sum_{i=1}^{N}s_{t,i}h_i^e \tag{7}$$

c_t 是所有编码状态与对应注意力大小乘积的加权平均。将 c_t 与 h_t^d 拼接并使用 tanh 函数激活，得到 t 时刻注意力向量为

$$a_t=\tanh(cat[c_t,h_t^d]) \tag{8}$$

由于每个时刻生成的词是从众多词中进行选择的，属于多分类问题，因此使用了 $soft\max$ 函数求出 t 时刻预测输出词 y_t 在词典中的概率：

$$P(y_t\mid y_{<t},X)=soft\max(a_t) \tag{9}$$

$$\Pr ob(y_t)=P(y_t\mid y_{<t},X) \tag{10}$$

2.3　复制机制

借鉴了文献[16]的算法思想，复制机制能有效解决 seq2seq 架构中输出长度固定的问题，其可根据输入序列对输

出序列进行动态调整。对于改进后的指针网络，整个词典中所有词的输出概率由两部分组成：a)源序列中每个词的注意力大小；b)网络使用固定目标词典自主生成每个词的概率大小。因此，将注意力值和生成概率相加，就得到最终每个词的输出概率，这相当于将原来固定的词典进行了动态扩展，所扩展的词典中包含了原来固在定词典中的词以及在源序列中的 OOV 词。具体的，假设 t 时刻输出 y_t 时所对应的复制和生成的软转换为 $P_{copy}(y_t)$，将注意力向量 a_t 输入到一个线性函数中，并使用 *sigmoid* 函数计算 $P_{copy}(y_t)$：

$$P_{copy}(y_t)=sigmoid(linear(a_t)) \tag{11}$$

$linear(\bullet)$ 表示线性函数，它的输出特征是一维向量；而 $P_{copy}(y_t)\in[0,1]$，它可以一次性提高或降低网络中所有生成词和复制词的概率。一方面，由网络自主生成 y_t 的概率为

$$P_{soft\max}(y_t)=\Pr ob(y_t)\times(1-P_{copy}(y_t)) \tag{12}$$

$\Pr ob(y_t)$ 由式(10)计算得到。另一方面，当一个词在源序列中出现多次时，则在该词对应词典索引的位置加上它在源序列出现的总的注意力值，所以使用注意力值作为 y_t 输出的概率为

$$P_{copy\_prob}(y_t)=(P_{copy}(y_t)\times\sum_{i:x_i=y_t}^{N}s_{t,i})\times map\_matrix \tag{13}$$

其中，map_matrix 是将源序列中的词映射到扩展词典中对应索引的矩阵(比如，源序列中词 x_i 的概率值是 0.1，则在扩展词典中 x_i 所对应的索引处加上 0.1)。所以，即使 x_i 不在原来的固定词典中，通过对固定词典的动态扩展，也能保证 x_i 的概率不为零，从而有效解决 OOV 词的问题。最后 y_t 在扩展词典中的概率由 $P_{soft\max}(y_t)$ 和 $P_{copy\_prob}(y_t)$ 两部分组成：

$$P(y_t)=P_{soft\max}(y_t)+P_{copy\_prob}(y_t) \tag{14}$$

若 $P_{soft\max}(y_t)$ 为零，表示 y_t 属于 OOV 词；而如果 $P_{copy\_prob}(y_t)$ 为零，表示 y_t 没有出现在源序列中。

2.4 input-feeding 方法

由于每个时刻的注意力决定都是相互独立的，在每个时刻只能使用当前时刻的注意力信息，导致模型在每个时刻的注意力决定的准确率降低。对于标准的序列生成模型来说，每个时刻的对齐决定应该考虑过去时刻的对齐信息，所以引入了文献[7]的 input-feeding 方法对上一时刻的注意力决定进行跟踪，以确保每个时刻的注意力能作出更加准确的决定。具体的，不将词嵌入向量直接输入到 LSTM 单元计算解码状态，而是将其与上一时刻的注意力向量拼接得到一个新的向量输入到 LSTM 单元中。假设 $t-1$ 时刻的注意力向量为 a_{t-1}，将 a_{t-1} 与输入的词嵌入向量 $emby_t$ 拼接，则得到新的向量 d_t：

$$d_t=cat[emby_t,a_{t-1}] \tag{15}$$

使用 d_t 代替式(4)中的 $emby_t$，最终得到新的 h_t^d 为

$$h_t^d=LSTM(h_{t-1}^d,d_t) \tag{16}$$

2.5 模型推断

对于已经训练好的模型，在解码推断时，输入测试文本 X^*，使用 beam-search 算法以最大化得分函数 $S(Y^*,X^*)$ 输出摘要 Y^*。然而，在解码推断中如果使用纯粹的 beam-search 算法将会导致输出序列的长度波动很小且都是输出较短的序列。原因是在每个解码时刻都添加了一个负对数概率值，对较长的句子会产生更低的分数，使得最终生成的摘要都倾向短序列，所以纯粹的 beam-search 算法对短序列来说是有利的，而对长序列是不利的。针对该问题，使用了长度规范化鼓励输出更长的序列，即除以生成摘要 Y^* 的长度来进行长度惩罚处理，将得分函数定义为

$$S(Y^*,X^*)=\frac{\log(P(Y^*\mid X^*))}{len(Y^*)} \tag{17}$$

式(17)中，$len(Y^*)$ 是循环生成的摘要 Y^* 的长度。这里使用了 Wu 等人提出的经验公式[18]计算 $len(Y^*)$：

$$len(Y^*)=\frac{(5+|Y^*|)^{\alpha}}{(5+1)^{\alpha}} \tag{18}$$

其中，α ($\alpha\in[0,1.2]$)是模型推断中需要设定的参数，通过调整 α 的值，可控制输出更多样化的摘要。

3 实验

3.1 数据集和数据预处理

实验中使用了 NLPCC2018(http://tcci.ccf.org.cn/conference/2018/index.php)提供的单一文本摘要数据集，包含了今日头条新闻的内容和标题，训练集为 5 万条，验证集和测试集各为 2 千条，数据预处理方法如下：

a)删除文本中图片和视频的长链接，同时使用正则表达式除去了一些无用字符，将每条新闻和对应的摘要分别截取为 400 个字符和 100 个字符；

b)使用 jieba(https://pypi.org/project/jieba/)对文本进行分词，并以字符为单位对文本进行分割，如果某个词在词典中的概率为零，则使用 UNK 标记代替；

c)人工构建了含有 300 个专有名词的词典用于分词，以提高分词准确率。

3.2 模型的算法和参数设定以及训练

3.2.1 模型算法

通过对源序列的编码表示，解码器每个时刻解码生成一个词，log 条件概率表示为

$$\log(P(Y\mid X))=\sum_{t=1}^{T}\log(P(y_t\mid y_{<t},X)) \tag{19}$$

因此，对于源序列(X)和目标序列(Y)一一对应的数据集 D，实验中训练的目标是最小化交叉熵损失函数：

$$Cel=\sum_{(X,Y)\in D}-\log(P(Y\mid X)) \tag{20}$$

鉴于本文的复制机制和 input-feeding 方法是在带有注意力机制的 seq2seq 架构的基础上实现的，因此将本文方法融合到解码器中，并给出了算法实现。

算法 1　模型算法

输入：(X={ x_1 ,..., x_N },Y={ y_1 ,..., y_T }) $\in D$

输出：model's parameters

foreach (X,Y) $\in D$:

encoder:　　/*编码器*/

for (x_i) in X:

$embx_i$ =embedding(x_i)　　/*词嵌入层*/

h_t = BiLSTM (h_{t-1},$embx_i$)

decoder:　　/*解码器*/

for (y_t) in Y:

$emby_t$ = embedding(y_t)

d_t =cat([$emby_t$, a_{t-1}])　　/*input-feeding 方法*/

h_t = LSTM (h_{t-1},d_t)

$a_t,s_{t,i}$ = attention(h_t , h_{i-1})　　/*注意力层*/

/*复制机制*/

$P_{prob}(y_t)=soft\max(a_t)$

$P_{copy}(y_t)=sigmoid(linear(a_t))$

$P_{soft\max}(y_t)=P_{prob}(y_t)\times(1-P_{copy}(y_t))$

$P_{prob\_copy}(y_t)=P_{copy}(y_t)\times s_{t,i}\times map\_matrix$

$P_{final}(y_t)=P_{soft\max}(y_t)+P_{prob\_copy}(y_t)$

$output(y_t)=\log(P_{final}(y_t))$

compute loss

update model's parameters

3.2.2 参数设定

根据文献[11]的研究结论，将基于词(word)和字符(char)两种级别的标记作为网络输入，词嵌入向量在网络中随机初

始化，通过网络自主学习，并采用 Adagrad 算法[19]对网络参数进行优化。模型的具体超参数设定如表 1 所示。

表 1　模型超参数设定

Table 1　Model's hyperparameters setting

超参数类型	超参数值
batch size	4
hidden state dimension	512
word embedding	128
learning rate	0.15
training steps	100000

3.2.3 模型训练

本次实验平台为 Ubuntu 16.04，内存 8 GB，NVIDIA GTX 1060 GPU，程序采用 Python 3.6 并基于 Pytorch (https://pytorch.org/)计算框架实现。模型中的所有参数均在 LSTM 单元中更新，训练时间约为 5.5 h。经过多次实验尝试，在模型推断时，设定 beam-size=10，α=1.2。图 3 为 char 模型训练时正确率的变化情况。

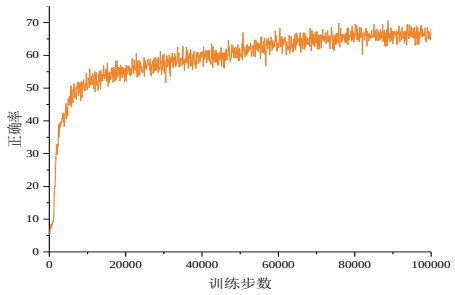

图 3　char 模型训练的正确率波动曲线

Fig. 3　Accuracy rate fluctuation curve in char model training

最初，随着训练步数的增加，正确率也在上升，当训练步数再增加时，正确率的变化很小，说明模型训练已经达到收敛。同时，因为输出的摘要还需尽可能保持流畅，所以在训练模型的过程中还需要关注的一项重要指标是困惑度(perplexity，ppl)，其是用于衡量语言概率模型优劣的一种方法，句子 Y 的困惑度定义为

$$ppl(Y)=\sqrt[T]{1/(P(y_1,y_1,...,y_T))} \tag{21}$$

由式(21)可知，当整个句子的概率值越大，得到的 *ppl* 值越小，则对应的语言模型越好。图 4 为训练过程中 *ppl* 值的变化率，当训练结束时，*ppl* 处于一个较小的值，说明训练得到了一个较好的的语言模型。

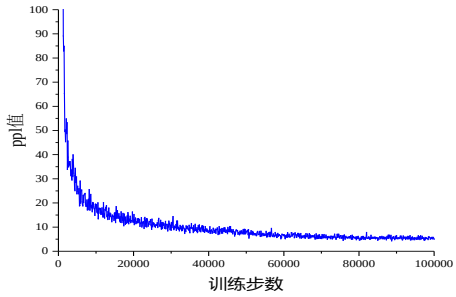

图 4　char 模型训练的 *ppl* 值波动曲线

Fig. 4　*Ppl* fluctuation curve in char model training

3.3　实验结果与分析

由 Lin 等人提出的 ROUGE (recall-oriented understudy for gisting evaluation)[20]广泛应用于自动摘要评价中。实验采用 ROUGE-1(R-1)、ROUGE-2(R-2)和 ROUGE-L(R-L)作为评价标准，分别表示要评估摘要和标准摘要之间的单个词重叠、两个词重叠和最长公共序列的比例。以下两部分内容分别对本文所提出模型(称为 seq2seq+attn+copy_if)与抽取式方法和生成式方法的实验结果(取 F 值)进行了分析。其中，表 2~4 的“(* word/char)”表示词典中的 word 或 char 的个数。

3.3.1 与抽取式方法比较的实验结果分析

表 2 列出了 seq2seq+attn+copy_IF 模型与经典的抽取式方法的实验比较结果。其中，MMR[2]是最大边际相关方法；TextRank[2]是基于图排序的方法；LexRank[2]是以特征向量为中心给句子排序的方法。

表 2　抽取式方法对比结果　/%

Table 2　Extractive methods comparison result

模型	R-1	R-2	R-L
MMR	22.50	12.12	17.23
LexRank	25.68	14.98	18.18
TextRank	28.57	17.43	18.56
seq2seq+attn+copy_if(6000 char)	**39.12**	**25.13**	**31.89**

通过表 2 可见，基于生成式方法的模型明显优于抽取式方法的模型。因为抽取式模型是基于统计学的方法来进行建模，这种方法难以捕获更深层的语义信息，使得词与词之间以及上下文之间没有语义上的联系；同时，因为是以句子为单位抽取摘要，所以抽取得到的句子中含有许多与主题无关的非关键词，这也会影响模型的准确率。而对于生成式方法而言，由于网络具有学习特征表示的能力，能更好的为每个词之间建立语义层面的联系，从而增强了上下文的联系；且生成摘要能对主题信息进行更精确的表达，减少了冗余信息，从而提高了 ROUGE 得分。

3.3.2 与生成式方法比较的实验结果分析

表 3 和 4 分别是 seq2seq+attn+copy_IF 的 word 级别和 char 级别模型与生成式方法模型的比较结果。其中 seq2seq+attn[11]为注意力机制的 seq2seq 模型；seq2seq+attn_IF 为融入了 input-feeding 方法的 seq2seq+attn 模型；pointer-generator[16]为具有指针网络的 seq2seq+attn 模型。

表 3　生成式方法对比结果(word)　/%

Table 3　Abstractive methods comparison result(word)　/%

模型	R-1	R-2	R-L
seq2seq+attn(50000 word)	18.76	8.49	16.38
seq2seq+attn(100000 word)	15.77	7.04	13.59
seq2seq+attn_IF(50000 word)	19.66	8.55	17.18
seq2seq+attn_IF(100000 word)	17.63	7.59	15.45
pointer-generator(50000 word)	25.32	11.76	21.57
seq2seq+attn+copy_if(50000 word)	**28.28**	**13.80**	**24.06**
seq2seq+attn+copy_if(100000 word)	**28.52**	**13.94**	**24.22**

表 4　生成式方法对比结果(char)　/%

Table 4　Abstractive methods comparison result(char)　/%

模型	R-1	R-2	R-L
seq2seq+attn(3000 char)	35.11	21.64	28.94
se2seq+attn(6000 char)	34.62	21.54	28.75
seq2seq+attn_IF(3000char)	36.14	22.32	29.03
seq2seq+attn_IF(6000 char)	35.96	22.18	28.84
pointer-generator(6000 char)	36.90	23.30	30.87
seq2seq+attn+copy_if(3000 char)	**39.03**	**24.93**	**31.51**
seq2seq+attn+copy_if(6000 char)	**39.12**	**25.13**	**31.89**

a) seq2seq+attn、seq2seq+attn_IF 的 word 和 char 级别模型希望通过增加词典中词的个数来提升 ROUGE 得分，但是从表 3 和 4 发现，在增加词个数的情况下，ROUGE 得分没有得到提升，甚至还出现了下降。而 seq2seq+attn+copy_if 模型在不同大小词典下的 ROUGE 得分也较为接近。由此说明，通过扩大词典的方法对最终结果的影响较小，从而验证了使用复制机制可以有效提高 ROUGE 得分。

b) seq2seq+attn+copy_if 模型与现阶段最优模型之一的

pointer-generator 模型相比，word 级别和 char 级别的结果都得到了提高，表明了 input-feeding 方法能在一定程度上跟踪历史时刻的注意力决定信息，使网络能学习到更多已生成序列的历史信息，从而提高了各个时刻的注意力决定的准确率；

c) 由表 3 和 4 还可得知，char 级别的结果明显优于 word 级别的结果。一种原因是数据集中的一些低频词没有登录到分词工具的语料库中，导致未登录词没有得到正确分割，使得输出的摘要中含有较多 UNK，而 char 级别的模型可在一定程度上解决未登录词的问题。

在实验中还发现，输出摘要的长短对 ROUGE 得分具有一定影响，图 4 展示了 char 级别模型在不同输出长度下 ROUGE 得分均值的变化情况。从图 5 可见，当输出字符个数在一定范围内波动时，对 ROUGE 得分的影响较小。对于本实验中的所有生成式模型而言，如果希望通过增加输出字符个数来提升 ROUGE 得分，图 4 证明了这种方法没有效果，因为盲目增加字符个数，将会导致许多与标准摘要无关的字符生成，由文献[20]可知，长度增加的情况下，如果含有的关键词减少，则会使 ROUGE 得分降低。

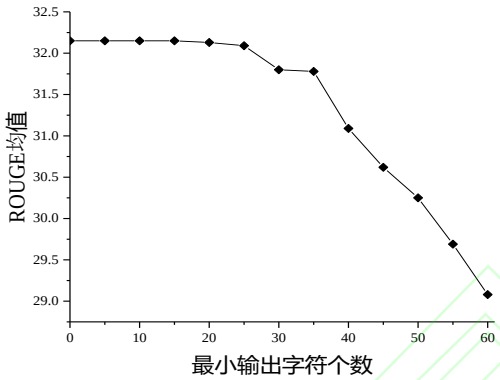

图 5　char 级别模型输出字符个数对 ROUGE 均值的影响

Fig. 5　The effect of ROUGE score mean in the char's model outputing character numbers

融合复制机制和 input-feeding 方法的自动摘要模型的 R-2 和 R-L 值提高反映了摘要的流畅性也得到了提高。图 6 列出了 seq2seq+attn+copy_if(char)模型输出摘要的示例。从输出的结果上看，摘要具有较强的可读性。

Article:大神爆料北京时间 7 月 7 日消息，据《雅虎体育》报道，爆料王阿德里安-沃纳罗斯基称，莫里斯-威廉姆斯已经与克里夫兰骑士达成了签约协议，他将签下一份 2 年 430 万美元的合同…… **BenchMark:**据雅虎体育，莫里斯-威廉姆斯与骑士达成协议，将与詹皇重聚，他将签下 2 年 430 万美元的合同 **seq2seq+attn+Copy_IF(char):**《雅虎体育》报道，莫里斯-威廉姆斯已与克里夫兰骑士达成签约协议，2 年 430 万美元的合同
Article:德国脚双星德拉克斯勒与格策阿森纳新赛季英超首轮比赛输球，这让他们的冲冠梦遭遇重击，同时也让温格明白，自己的球队仍然需要补充实力。据英国《每日电讯报》透露，温格在收购本泽马遇挫之后，将豪砸 8600 万英镑收购格策和德拉克斯勒。阿森纳输给西汉姆联之后，温格承诺，如果市场上有实力出色的球员…… **BenchMark:**据外媒报道，温格在收购本泽马遇挫之后，将豪砸 8600 万英镑收购格策和德拉克斯勒 **seq2seq+attn+Copy_IF(char):**温格在收购本泽马遇挫之后，将豪砸 8600 万英镑收购格策和德拉克斯勒

图 6　seq2seq+attn+copy_if(char)模型输出摘要示例

Fig. 6　Summary examples by seq2seq+attn+copy_if generated

4　结束语

通过以编码器-解码器为基础，使用 GlobalAttention 计算注意力向量和注意力分布，接着在解码器中融合了复制机制和 input-feeding 方法提出了面向中文的自动摘要模型，有效解决了 OOV 词的问题，并改善了注意力决定性能，最终提高了中文自动摘要的准确率。

由于数据集较小，模型容易出现过拟合现象，导致模型在训练和测试过程中存在有一定误差。近年来，迁移学习在自然语言处理领域受到高度关注。通过将预训练好的语言模型或词嵌入模型迁移到特定任务中，不但能提升模型的训练效率，还能有效解决数据量少的问题，在使用少量数据的情况下仍能提升模型性能。同时，因为理论算法的最终目的是将训练得到的模型应用于实际中，才能体现出算法研究的价值，将来的工作之一是将本文模型部署到实际应用中，以实现从理论研究到实际应用的有效衔接。因此，下一步工作将这两部分作为重点，在提高中文自动摘要质量的基础上并将模型部署到实际应用中。

参考文献：

[1] Allahyari M, Pouriyeh S, Assefi M, *et al*. Text summarization techniques: a brief survey [J]. International Journal of Advanced Computer Science & Applications, 2017, 8 (10): 397-405.

[2] Gambhir M, Gupta V. Recent automatic text summarization techniques: a survey [J]. Artificial InTelligence Review, 2017, 47 (1): 1-66.

[3] 余珊珊，苏锦钿，李鹏飞. 基于改进的 TextRank 的自动摘要提取方法 [J]. 计算机科学，2016，43 (6): 240-247. (Yu Shanshan, Su Jindian, Li Pengfei. Improved TextRank-based method for automatic summarization [J]. Computer Science, 2016, 43 (6): 240-247.)

[4] 刘彼洋，孙锐，姬东鸿. 基于矩阵分解和子模最大化的微博新闻摘要方法 [J]. 计算机应用研究，2017，34(10): 2892-2896，2928. (Liu Biyang, Sun Rui, Ji Donghong. Weibo-oriented news summarization based on matrix factorization and submodular maximization [J]. Application Research of Computers, 2017, 34(10): 2892-2896, 2928.)

[5] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks [C]// Advances in Neural Information Processing Systems. Montreal: NIPS Press, 2014: 3104-3112.

[6] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [EB/OL]. (2014) [2019-03-01]. http://arxiv. org/abs/1409. 0473.

[7] Luong T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation [C]// Proc of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: EMNLP Press, 2015: 1412-1421.

[8] Mou Lili, Song Yiping, Yan Rui, *et al*. Sequence to backwardand forward sequences: a content-introducing approach to generative short-text conversation [C]// The 26th International Conference on Computational Linguistics. Osaka: Coling Press, 2016: 3349-3358.

[9] Rush A M, Chopra S, Weston J. A neural attention model for abstractive sentence summarization [C]// Proc of the 2015 on Empirical Methods in Natural Language Processing. Lisbon: EMNLP Press, 2015: 577-585.

[10] Chopra S, Auli M, and Rush A M. Abstractive sentence summarization with attentive recurrent neural networks [C]// Proc of the 15th Conference of the North American Chapter of the Association for Computational Linguistics. San Diego: ACL Press, 2016: 93-98.

[11] Hu Baotian, Chen Qingcai, Zhu Fangze. LCSTS: A large scale chiese short text summarization dataset [C]// Proc of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: EMNLP Press, 2015: 1967-1972.

[12] Li Piji, Lam W, Bing Lidong, *et al*. Deep recurrent generative decoder for abstractive text summarization [C]// Proc of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen: EMNLP Press, 2017: 2091-2100.

[13] Oriol V, Meire F, and Navdeep J. Pointer networks [C]// Advances in

Neural Information Processing System. Montreal: NIPS Press, 2015: 2692-2700.

[14] Gu Jiatao, Lu Zhengdong, Li Hang, *et al.* Incorporating copying mechanism in sequence-to-sequence learning [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: ACL Press, 2016: 1631-1640.

[15] 应文豪，李素建，穗志方．一种话题敏感的抽取式多文档摘要方法 [J]．中文信息学报，2017，31 (6)：155-161. (Ying Wenhao, Li Sujian, Sui Zhifang. A topic-sensitive extractive method for multi-document summarization [J]. Journal of ChineseInformation Processing, 2017, 31 (6): 155-161.)

[16] See A, Liu P J, Manning, C D. Get to the point: summarization with pointer-generator networks [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Vancouver: ACL Press, 2017: 1073-1083.

[17] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation. 1997, 9 (8): 1735-1780.

[18] Wu Yonghui, Schuster M, Chen Zhifeng, *et al.* Google's neural machine translation system: bridging the gap between human and machine translation [EB/OL]. (2016) [2019-03-01]. https://arxiv. org/abs/1609. 08144.

[19] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization [J]. Journal of Machine Learning Research. 2011, 12 (Jul): 2121-2159.

[20] Lin C Y, Hovy E. Automatic evaluation of summaries using N-gram co-occurrence statistics [C]// Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology. Sapporo: ACL Press, 2003: 71-78.