

融合词汇特征的生成式摘要模型

江跃华¹, 丁磊², 李娇娥¹, 杜皓暄³, 高凯¹

(1.河北科技大学信息科学与工程学院,河北石家庄 050018;2.石家庄市公安局信息中心,河北石家庄 050021;3.西安电子科技大学通信工程学院,陕西西安 710126)

摘要:为了能在摘要生成过程中利用词汇特征(包含 n-gram 和词性信息)识别更多重点词汇内容,进一步提高摘要生成质量,提出了一种基于 sequence-to-sequence(Seq2Seq)结构和 attention 机制的、融合了词汇特征的生成式摘要算法。算法的输入层将词性向量与词向量合并后作为编码器层的输入,编码器层由双向 LSTM 组成,上下文向量由编码器的输出和卷积神经网络提取的词汇特征向量构成。模型中的卷积神经网络层控制词汇信息,双向 LSTM 控制句子信息,解码器层使用单向 LSTM 为上下文向量解码并生成摘要。实验结果显示,在公开数据集和自采数据集上,融合词汇特征的摘要生成模型性能优于对比模型,在公开数据集上的 ROUGE-1, ROUGE-2, ROUGE-L 分数分别提升了 0.024, 0.033, 0.030。因此,摘要的生成不仅与文章的语义、主题等特征相关,也与词汇特征相关,所提出的模型在融合关键信息的生成式摘要研究中具有一定的参考价值。

关键词:自然语言处理;文本摘要;注意力机制;LSTM;CNN

中图分类号:TP319 **文献标志码:**A

Abstractive summarization model considering hybrid lexical features

JIANG Yuehua¹, DING Lei², LI Jiaoe¹, DU Haoxuan³, GAO Kai¹

(1.School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China;2.Information Center of Shijiazhuang Public Security Bureau, Shijiazhuang, Hebei 050021, China;3. Xi'dian University, Xi'an, Shaanxi 710126, China)

Abstract: In order to use lexical features (including n-gram and part of speech information) to identify more key vocabulary content in the summarization generation process to further improve the quality of the summarization, an algorithm based on sequence-to-sequence (Seq2Seq) structure and attention mechanism and combining lexical features is proposed. The input layer of the algorithm combines the part of speech vector with the word vector, which is the input of the encoder layer. The encoder layer is composed of bi-directional LSTM, and the context vector is composed of the output of the encoder and the lexical feature vector extracted from the convolution neural network. The convolutional neural network layer in the model controls the lexical information, the bi-directional LSTM controls the sentence information, and the decoder layer uses unidirectional LSTM

收稿日期:2018-10-14;修回日期:2019-03-01;责任编辑:陈书欣

基金项目:国家自然科学基金(61772075);河北省自然科学基金(F2017208012);教育部人文社会科学研究专项任务项目(工程科技人才培养研究)(17JDGC022)

第一作者简介:江跃华(1993—),男,河北邯郸人,硕士研究生,主要从事自然语言处理方面的分析研究。

通信作者:高凯教授。E-mail:gaokai@hebust.edu.cn

江跃华,丁磊,李娇娥,等.融合词汇特征的生成式摘要模型[J].河北科技大学学报,2019,40(2):152-158.

JIANG Yuehua, DING Lei, LI Jiaoe, et al. Abstractive summarization model considering hybrid lexical features[J]. Journal of Hebei University of Science and Technology, 2019, 40(2): 152-158.

to decode the context vector and generates the summarization. The experiments on public dataset and the self-collected dataset show that the performance of the summarization generation model considering lexical feature is better than that of the contrast model. The ROUGE-1, ROUGE-2 and ROUGE-L scores on the public dataset are improved by 0.024, 0.033 and 0.030, respectively. Therefore, the generation of summarization is not only related to the semantics and themes of the article, but also to the lexical features. The proposed model provides a certain reference value in the research of generating summarization of integrating key information.

Keywords: natural language processing; text summarization; attention mechanism; LSTM; CNN

近年来,由于互联网技术的迅猛发展以及用户规模的爆发式增长,互联网的数据呈现指数级增长。文本是目前形式多样的信息数据中最主要的存在形式。如何能从海量的文本数据中挖掘出重要而且是用户关心的话题和信息,对于提升阅读效率、消化海量信息是非常有帮助的。文本摘要是文本信息自动抽取的主要任务之一,它追求以更加精简、精确的方式,用少量文本尽可能地表述原文含义。按照产生方式的不同,文本自动摘要可分为抽取式和生成式摘要。其中,抽取式摘要主要是指从原文中抽取和文章主题相关的内容,在不超过一定字数的限制下,尽可能地让抽取的句子覆盖到原文所有含义,并按顺序排列生成一段摘要文本;生成式摘要则是通过让模型学习原文中所表达的含义,推断所需要生成的摘要内容,由模型自动生成一段摘要文本。根据任务的不同,文本摘要也可分单文档摘要和多文档摘要。通常,人们在阅读文章时,一方面要保留句子的大部分信息,另一方面要关注句子中重要的词汇信息。在生成摘要时,应对原文中的短语按照词性做不同的处理,例如对名词、动词等有实际意义的词汇信息,在生成摘要时保留信息;对于介词、虚词等无明显含义的词汇应保留较少的信息。

在 Seq2Seq 框架和 attention 机制的基础之上提出了使用卷积神经网络提取词汇特征,并将其融合在上下文中的方法,对原文的词汇特征进行提取,并和基础框架所提取的句子特征一起构成上下文特征矩阵,目的在于使用上下文信息时既要保留文本的序列信息,同时也要保留重点词汇特征信息。

1 相关工作

在生成单文档摘要时,可通过深度学习的方式实现,即使用大规模的数据集,以数据驱动的方式训练神经网络,学习原文浅层语义并生成摘要。在相关工作中,RUSH 等^[1]和 CHOPRA 等^[2]使用 Seq2Seq 框架,在编码器端给定输入的句子,在解码器端输出摘要的每个词,并以局部注意力机制提升效果。NALLAPATI 等^[3]融入传统的 TF-IDF, POS、命名实体等特征来提升效果,引入 Pointer 解决未知词和低频词。HU 等^[4]提供了一个新的中文短文本摘要数据集(数据来源为新浪微博),并根据正文和摘要之间的相关性进行了人工评分。YAO 等^[5]总结了近几年文本摘要领域的新方法,从抽取式和生成式、单文档和多文档这几方面进行评价,包括数据集和评价方法等。摘要生成时,其输出句子的长度较难把控,为解决这个问题, KIKUCHI 等^[6]提出在解码器端,将原摘要的长度信息输入到 LSTM 中进行控制,取得了较好的结果。ZENG 等^[7]考虑到人工编写摘要时通常需要先对全文内容通读以了解文章主题,提出将全文信息通过 RNN 编码为一个向量,再进行解码。SEE 等^[8]通过在输出端对下一个词的生成或复制的概率进行预测,解决 OOV(out-of-vocabulary)和低频词的问题。GEHRING 等^[9]将编码器和解码器用卷积神经网络替代,达到了接近 state-of-the-art 的表现。CHANG 等^[10]在生成中文摘要时,当编码器输入词时可使模型获得较多的信息,解码器输出字时可压缩并精简文章的内容。MA 等^[11]在训练时将摘要内容进行自编码,监督生成的原始内容编码,使编码器生成的内容更加接近参考摘要。PAULUS 等^[12]提出内注意力机制解决重复词语的问题,训练方法中融合了监督学习和强化学习以提升效果。WANG 等^[13]以 convolutional sequence to sequence(ConvS2S)为基础,将主题信息融入其中,使用强化学习训练方法优化。FAN 等^[14]根据用户写作摘要的风格,包括摘要长度、用词等特征生成摘要。GAO 等^[15]提出了读者感知生成摘要的任务,设计了一个名为读者感知摘要生成器(RASG)的对抗性学习模型,将读者的评论纳入生成式摘要。针对 PECH(产品感知答案生成)任务,GAO 等^[16]设计了一种基于对抗性学习的模型 PAAG,提出一种基于注意力机制的阅读审查方式,用来捕获给定的问题中与评论相关度最高的词语。

和上述工作不同的是,笔者提出基于 Seq2Seq 的框架,在上下文特征中加入由卷积神经网络提取的词汇特征,根据词性信息判断词汇的重要性,进而保留文章的重点内容,完成文本摘要的生成。

2 融合词汇特征模型

以 Seq2Seq 为基础框架模型,结构如图 1 所示。在输入层,使用词性向量和词向量进行叠加,构成整个网络的输入向量。在编码器端,分为 2 部分:一部分使用双向 LSTM^[17] 模型,对分词后的源文本内容进行句子级别的浅层表征学习;另一部分使用卷积神经网络对所有词汇提取 n-gram 和词性特征,最后将 2 部分学习到的特征矩阵融合在一起,构成上下文向量。在输出摘要时使用单向的 LSTM,每个时间步都需要对上下文特征矩阵进行注意力权重的分配,进而生成词汇。

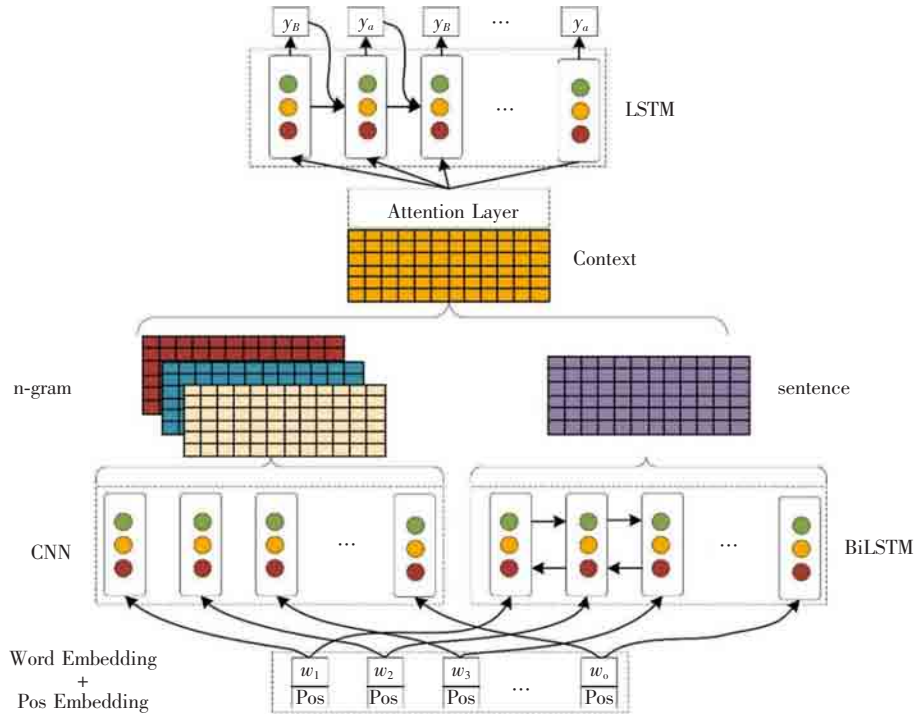


图 1 融合短语特征的摘要生成模型

Fig.1 Abstract generation model of hybrid phrase features

2.1 构建输入层向量

融合短语特征的自动摘要模型的输入由 2 部分构成,第 1 部分将分词后的源文本转换成词向量,使用 $\mathbf{x}^w = \{x_1^w, x_2^w, \dots, x_n^w\}$ 来表示;第 2 部分将源文本的词性转换为向量,用 $\mathbf{x}^p = \{x_1^p, x_2^p, \dots, x_n^p\}$ 表示,其中 n 表示输入词性的索引。最后,对输入序列的词向量和词性向量连接,得到带有词性特征的输入向量 \mathbf{w}^p ,如式(1)所示。

$$\mathbf{w}_n^p = \sum_{i=1}^n [x_i^w ; x_i^p]. \tag{1}$$

2.2 Seq2Seq 模型

编码器使用双向 LSTM 模型,输入的是带有词性特征的向量,通常将编码器所有输出的隐藏状态作为原始句子的浅层语义表示,在解码器端对该向量进行解码。由于模型所使用的是双向 LSTM,每个时间步都接收来自前一时刻或后一时刻的隐层状态,所以在每个时间步需要对前向和后向的各个隐层状态向量进行连接,得到包含前后语义的隐藏状态。式(2)中 \vec{c}_t, \vec{h}_{t+1} 表示后一时刻的隐层向量,式(3)中 \vec{c}_t, \vec{h}_{t-1} 表示前一时刻的隐层向量,式(4)中 \vec{h}_t 表示当前时刻融合前向和后向的隐层向量,对于解码器同样使用 LSTM 神经网络进行解码。

$$\vec{c}_t, \vec{h}_t = \text{LSTM}(w_t^p, \vec{c}_{t+1}, \vec{h}_{t+1}), \tag{2}$$

$$\vec{c}_t, \vec{h}_t = \text{LSTM}(w_t^p, \vec{c}_{t-1}, \vec{h}_{t-1}), \tag{3}$$

$$\vec{h}_t = [\vec{h}_t; \vec{h}_t]. \tag{4}$$

式(5)中 s_t 表示解码器端的隐藏向量, c_t 表示 LSTM 中的状态向量。式(6)中 y_t 表示每个时间步输出的摘要序列词汇。

$$\mathbf{s}_t = \text{LSTM}(\mathbf{c}_t, \mathbf{y}_{t-1}, \mathbf{s}_{t-1}), \quad (5)$$

$$\mathbf{y}_t = \text{softmax}(g[\mathbf{c}_t; \mathbf{s}_t]). \quad (6)$$

2.3 融合词汇特征

借鉴 LIN 等^[18]的工作内容,依据其设定的卷积神经网络结构,卷积核的大小依次设定为 $k=1, k=3$,使卷积神经网络能提取词汇的 n -gram 的特征。在卷积层,将包含词性的词性序列作为基本单位输入到网络中,使用多个与输入向量维度相一致的卷积单元学习词汇特征。假设整个输入序列的长度为 n ,对于每个卷积单元,经过一次卷积后,生成与输入矩阵大小相同的特征矩阵,最后将多个卷积结果和编码器的隐层状态连接,使用全连接层学习多种特征融合。

$$\mathbf{h} = \partial(W[\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3, \tilde{\mathbf{h}}] + \mathbf{b}). \quad (7)$$

式(7)中 \mathbf{h} 为最终的编码器隐层向量,维度为 $m \times n$, m 代表输入向量的维度, n 表示输入序列的长度,其中 $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$ 表示由卷积层输出的特征矩阵。 ∂ 表示激活函数 GLU^[19]。得到融合词汇特征和句子特征的隐层向量后,引入注意力机制^[20]来捕获输出内容与上下文向量的关联程度。

$$\mathbf{c}_t = \sum_{i=1}^N \alpha_i \mathbf{h}_i, \quad (8)$$

$$\alpha_i = \frac{\exp(e_{ij})}{\sum_{k=1}^T \exp(e_{ik})}, \quad (9)$$

$$e_{ij} = \alpha(\mathbf{s}_{i-1}, \mathbf{h}_j). \quad (10)$$

式(8)一式(10)中, \mathbf{c}_t 表示当前时刻的上下文向量, e_{ij} 表示解码器端隐层状态 \mathbf{s}_{i-1} 和编码器隐层状态 \mathbf{h}_j 之间的相关系数。当输出摘要时,需要将得到上下文的向量 \mathbf{c}_t 输入到解码器的 LSTM 单元中。

2.4 损失函数

本文训练的模型采用反向传播方式,目标函数(代价函数)为交叉熵代价函数,将 \mathbf{y}_t 作为输出生成的摘要词汇, X 为源文本输入序列,模型训练的目标是在给定输入语句的情况下最大化每个输出摘要词的概率。式(11)中 k 表示同一个训练批次句子的索引, t 表示句中输入词汇的索引。

$$\Delta = -\frac{1}{N} \sum_{k=1}^N \sum_{t=1}^T \log[p(\mathbf{y}_t^{(k)} | \mathbf{y}_{<t}^{(k)}, X^{(k)})]. \quad (11)$$

3 实验

3.1 数据集描述

本文模型所使用的第一个数据集是大规模的中文短文本摘要数据集 LCSTS^[4],主要是收集自新浪微博的内容,总共包含超过 240 多万条摘要文本内容。每对摘要文本中,其原始内容少于 140 个字,摘要则由人工编写。数据集共分为 3 部分,第 1 部分约 240 万对,第 2 部分约 1 万对,第 3 部分约 1 千对,其中第 2 部分和第 3 部分由人工方式对标题和正文之间的相关度按照 1 至 5 进行评分,得分越高相关度越高。按照 HU 等^[4]的工作内容,将数据集分割为训练集、验证集和测试集。另外,还使用第 2 个数据集,以人工方式收集并整理 20 条微博内容供模型测试使用,参考摘要和原文内容的抽取方式与 LCSTS^[4]数据集的抽取方式相同。

3.2 实验参数设置

模型中词向量的维度为 512,词性向量的维度为 50,经过网络合并输出后的特征维度为 562。以每 64 个样本为一个批次的方式进行模型训练,设置 Adam^[21]的学习率为 0.001,设置代价函数的惩罚项为 0.001。同时设置编码器的 LSTM 神经网络层数为 3 层,解码器的 LSTM 神经网络为 1 层,每个卷积层的输出维度与输入向量维度保持一致。该摘要生成模型运行在 Ubuntu 16.04 系统,运行环境为 PyTorch0.4 和 NVIDIA GTX 1080ti 显卡。

3.3 评价指标设定

实验借鉴 RUSH 等^[1]的工作内容,使用自动文本摘要的相关评测标准 ROUGE^[22]。该评测方法基于生成摘要和参考摘要中 n -gram 的共现信息来评价摘要质量,现在被广泛应用于 DUC 评测任务。评测标准包括 ROUGE-N, ROUGE-L 等。其中 N 表示 N 元词,而 L 表示最长公共序列。ROUGE-N 的计算方式为

$$\text{ROUGE-N} = \frac{\sum_{S \in \text{sumref}} \sum_{\text{gram}_n \in S} \text{count}_{\text{match}}(\text{gram}_n)}{\sum_{S \in \text{sumref}} \sum_{\text{gram}_n \in S} \text{count}(\text{gram}_n)}, \quad (12)$$

式中: sum_{ref} 表示参考摘要; $\text{count}_{\text{match}}(\text{gram}_n)$ 表示由模型生成的摘要和标准参考摘要中共同出现的 n 元词个数, $\text{count}(\text{gram}_n)$ 则表示标准参考摘要中出现 n 元词的个数。ROUGE-L 表示模型生成摘要与标准参考摘要之间,最长的公共序列长度与参考摘要长度之间的比值。ROUGE-L 相比于 ROUGE-N 考虑了摘要中词语的次序,评价更为合理。

4 实验结果及分析

本文所使用的对比模型是 RNN^[14]和 RNN content^[14]。其中,RNN content 使用了上下文向量 Content,将所有的编码器输出作为解码器的输入;RNN(W)使用的是 jieba 分词后的文本进行训练;RNN(C)使用基于词的方法训练模型。Seq2Seq 是本文所实现的基于 Seq2Seq 并结合注意力机制的模型,其编码器和解码器使用的分别为双向 LSTM 和单向 LSTM。

表 1 给出的是不同模型的 ROUGE 测度值。本文所提出的融合词汇特征的摘要生成模型与其他 4 种模型相比,性能有所提升。从表 1 中可以看出,使用双向 LSTM 作为解码器的 Seq2Seq 模型,要比单向的 RNN 作为编码器的模型在 ROUGE-1,ROUGE-2 和 ROUGE-L 上分别提升 0.013,0.020,0.019,说明使用双向 LSTM 更能捕捉文章特征信息。融合词汇特征的模型(HN)在基于 Seq2Seq 框架的基础上,增加词汇特征的融合要比不使用词汇特征的模型在 ROUGE-1,ROUGE-2 和 ROUGE-L 上的指标均有所提升。实验效果证明融合词汇特征的模型比无词汇特征的模型效果要更好一些。

表 1 在测试集 1 上 ROUGE 分数

Tab.1 ROUGE score in test set 1

模型	ROUGE-1	ROUGE-2	ROUGE-L
RNN(W)	0.177	0.085	0.158
RNN(C)	0.215	0.089	0.186
RNN content(W)	0.268	0.161	0.241
RNN content(C)	0.299	0.174	0.272
Seq2Seq	0.312	0.194	0.291
HN(hybrid n-gram)	0.323	0.207	0.302

本文选取 LCSTS 摘要数据集的训练集对融合词汇特征的摘要模型进行训练。采用其测试样例进行测试,从实验结果中抽取出部分样例如表 2 所示。从样例中可以看到,HN 模型生成的摘要内容更加接近参考摘要的内容,且对于生成的摘要来说,由于模型中融入了词汇特征,因而能保留原文的重点词汇信息。

表 2 测试集 1 上生成摘要样例对比

Tab.2 Summary on test set 1 generates the results

源文本:快递公司再现“暴力分拣”的丑闻。据中央电视台 24 日播放的画面,申通等快递公司的分拣工人还在包裹堆上随意行走;25 日,网上还传出了圆通快递“暴力分拣”的视频。截至昨天,申通快递已经承认错误,其他快递公司却保持沉默。
参考摘要:暴力分拣曝光后快递公司齐缄口。
Seq2Seq:“暴力分已经”曝光快递受质疑。
HN 模型:“暴力分拣”曝光后快递公司齐行走。

这里给出另一组与图 2 中的实验结果样例。源文本是:昨晚,南车、北车陆续发布公告。据一财记者多方了解,公告所提到的“筹划重大事项”,正是酝酿将南北车合并,合并一事由国务院要求推进,并由国务委员王勇负责督办,而合并背后的导火索,是两家公司在海外市场竞相压价的“恶性竞争”。生成的摘要内容:避免海外恶性竞争国务委员牵头推南北车合并。图 2 纵轴代表生成的摘要内容,横轴代表原文内容,由于篇幅所限,只截取具有代表性的内容。从图 2 中可以看到,在生成的摘要词汇和原文词汇间注意力权重的不同,如原文内容中的“合并一事由国务院要求推进”与摘要中的“国务委员牵头推南北车合并”,这两段文本间色块的深度明显大于其他部分,表示这两段内容间有较强的关联性。但图 2 中也有一些注意力权重未能达到

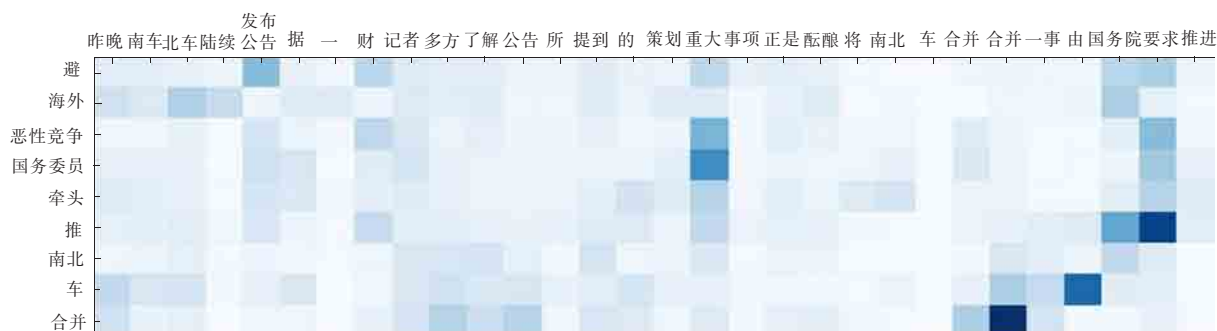


图 2 部分词汇的注意力权重分配

Fig.2 Distribution of attention weight for some words

预期效果,如原文中的“由国务院”的“由”字和摘要中“南北车”的“车”字相关性最高,导致权重分配错误的原因可能是因为在模型训练过程中,对一些频繁出现的搭配进行了错误的学习,解决途径之一是增加训练集,以降低某些错误固定搭配的比例。

表 3 所示的为本文所实现的 HN 模型和 Seq2Seq 模型在测试集 2 上的 ROUGE 分数。从表 3 中可看到 HN 模型生成摘要的 ROUGE 分数比 Seq2Seq 的要更高一点。说明本文所实现的模型在其他测试集上,也比单纯的结合注意力机制的 Seq2Seq 模型更优。但测试集 1 的 ROUGE 分数明显高于测试集 2 上的分数,主要由于测试集 2 为人工收集的数据集,考虑到成本问题,只有 20 条,数据不具有广泛的代表性,仅作为比较模型的优劣程度的指标。

表 3 在测试集 2 上的 ROUGE 分数

Tab.3 ROUGE score in test set 2

模型	ROUGE-1	ROUGE-2	ROUGE-L
Seq2Seq	0.274	0.148	0.257
HN(hybrid n-gram)	0.283	0.161	0.262

表 4 为模型在测试集 2 上的实验结果样例。

表 4 在测试集 2 上的摘要生成结果

Tab.4 Summary on test set 2 generates the results

源文本:16日凌晨,福建福州仓山区盖山镇叶厦村一栋砖混结构自建民房发生倒塌。当晚21点45分现场搜救工作完成,17名被困人员已陆续被救出,其中3名重伤者死亡。截至当晚23点,已有6人出院。目前警方已控制涉事房主,房屋倒塌原因正在调查。
参考摘要:福建福州一楼房倒塌事件搜救工作完成3死14伤。
Seq2Seq:福建房主一房屋倒塌已造成3人死亡。
HN模型:福建楼房倒塌已致3人死亡。

5 结 语

笔者提出的融合词汇特征的模型应用卷积神经网络从原文本中提取词汇特征,采用双向 LSTM 提取句子特征,然后将词汇特征与句子特征相融合,以达到利用词汇特征寻找文章中重点内容的需求。通过与其他模型相比较,证明了融合词汇特征的模型是有效的。

虽然所提出的模型达到了预期效果,但仍然有较大的提升空间,首先本文的模型只是通过非线性函数融合句子特征和词汇特征,未对词汇特征的权重进行进一步优化,以后可考虑使用多层次注意力机制,对词汇特征进一步提取。其次在做摘要任务时,需要考虑文章的主题信息,而且如果文章内容较长时,在模型的训练过程中通常会出现梯度爆炸或梯度消失等问题,也不利于模型提取重要信息。未来可进一步考虑在长文

本内容中,使用划分文章段落和句子的层次结构的方法缩短序列,达到分块提取重要信息的需求,同时可加入文章主题信息作为监督,保证模型生成的摘要和文章主题相一致。

参考文献/References:

- [1] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon:[s.n.], 2015:379-389.
- [2] CHOPRA S, AULI M, RUSH A M. Abstractive sentence summarization with attentive recurrent neural networks[C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. San Diego:[s.n.], 2016: 93-98.
- [3] NALLAPATI R, ZHOU B, SANTOS C N D, et al. Abstractive text summarization using sequence-to-sequence rnns and beyond[C]// Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning. Berlin:[s.n.],2016:280-290.
- [4] HU Baotian, CHEN Qingcai, ZHU Fangze. LCSTS: A large scale Chinese short text summarization dataset[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon:[s.n.],2015: 1967-1972.
- [5] YAO Jinge, WAN Xiaojun, XIAO Jianguo. Recent advances in document summarization[J]. Knowledge and Information Systems, 2017, 53(2): 297-336.
- [6] KIKUCHI Y, NEUBIG G, SASANO R, et al. Controlling output length in neural encoder-decoders[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin:[s.n.], 2016:1328-1338.
- [7] ZENG Wenyuan, LUO Wenjie, FIDLER S, et al. Efficient summarization with read-again and copy mechanism[C]// Proceedings of the International Conference on Learning Representations.[S.l.]:[s.n.],2017:1-13.
- [8] 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. [S.l.]:[s.n.],2017:1073-1083.
- [9] GEHRING J, AULI M, GRANGIER D, et al. Convolutional sequence to sequence learning[C]//Proceedings of the 34th International Conference on Machine Learning. [S.l.]:[s.n.],2017: 1243-1252.
- [10] CHANG C T, HUANG C C, HSU J Y J, et al. A hybrid word-character model for abstractive summarization[EB/OL]. <https://arxiv.org/pdf/1802.09968v2.pdf>, 2018-02-28.
- [11] MA Shuming, SUN Xu, LIN Junyang, et al. Autoencoder as assistant supervisor: Improving text representation for Chinese social media text summarization[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Short Papers). Melbourne:[s.n.], 2018:725-731.
- [12] PAULUS R, XIONG Caiming, SOCHER R. A deep reinforced model for abstractive summarization[C]//Proceedings of Sixth International Conference on Learning Representations. [S.l.]:[s.n.],2017:1-13.
- [13] WANG Li, YAO Junlin, TAO Yunzhe, et al. A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAD). [S.l.]:[s.n.], 2018: 4453-4460.
- [14] FAN A, GRANGIER D, AULI M. Controllable abstractive summarization[C]//Proceedings of the 2nd Workshop on Neural Machine Translation and Generation. Melbourne:[s.n.],2017:45-54.
- [15] GAO Shen, CHEN Xiuying, LI Piji, et al. Abstractive text summarization by incorporating reader comments[EB/OL]. <https://arxiv.org/pdf/1812.05407v1.pdf>, 2018-12-13.
- [16] GAO Shen, CHEN Xiuying, LI Piji, et al. Product-aware answer generation in e-commerce question-answering[C]//Proceedings of the 12th ACM International Conference on Web Search and Data Mining.[S.l.]:[s.n.],2019:07696.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [18] LIN Junyang, SUN Xu, MA Shuming, et al. Global encoding for abstractive summarization[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Short Papers). Melbourne:[s.n.], 2018:163-169.
- [19] DAUPHIN Y N, FAN A, AULI M, et al. Language modeling with gated convolutional networks[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney:[s.n.], 2017: 933-941.
- [20] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[C]// Proceedings of International Conference on Learning Representations.San Diego:[s.n.], 2015:1409. 0473V6.
- [21] KINGMA D P, BA J L. Adam: A method for stochastic optimization[C]//Proceedings of International Conference on Learning Representations.San Diego:[s.n.], 2015:1412.6980V9.
- [22] LIN C Y, HOVY E. Automatic evaluation of summaries using N-gram co-occurrence statistics[C]//Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology.Stroudsburg:[s.n.], 2003:71-78.