

基于分类的中文文本摘要方法

庞超 尹传环

(北京交通大学计算机与信息技术学院 北京 100044)

摘要 自动文本摘要是自然语言处理领域中一项重要的研究内容,根据实现方式的不同其分为摘录式和理解式,其中理解式文摘是基于不同的形式对原始文档的中心内容和概念的重新表示,生成的文摘中的词语无需与原始文档相同。提出了一种基于分类的理解式文摘模型。该模型将基于递归神经网络的编码-解码结构与分类结构相结合,并充分利用监督信息,从而获得更多的摘要特性;通过在编码-解码结构中使用注意力机制,模型能更精确地获取原文的中心内容。模型的两部分可以同时在大数据集下进行训练优化,训练过程简单且有效。所提模型表现出了优异的自动摘要性能。

关键词 递归神经网络,注意力机制,文本摘要,文本分类

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.01.024

Chinese Text Summarization Based on Classification

PANG Chao YIN Chuan-huan

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract Automatic text summarization is an important content in natural language processing. According to different implementation ways, it can be classified into extractive summarization and abstractive summarization. Abstractive summarization consists of ideas or concepts which are taken from the original document but are re-interpreted and shown in a different form, the aspects of which may not appear as part of the original document. This paper proposed an abstractive model with classifier. The model combines encoder-decoder structure based on recurrent neural networks with classifier to use supervised information more sufficiently and get more abstract features. However, encoder-decoder structure and classifier can easily be trained end-to-end and scale a large amount of training data at the same time. The model obtains good performance of text summarization and text classification.

Keywords Recurrent neural networks, Attention mechanism, Text summarization, Text classification

随着互联网的迅猛发展,人们越来越多地依赖互联网获取所需要的信息。如何在海量信息中寻找所需内容已经成为一个迫在眉睫的问题。利用自动文本摘要技术提取文本中的主要信息已成为解决该问题的重要手段之一。

文摘是可以准确全面地反映某一文献中心内容的简洁连贯的短文,自动文本摘要技术是利用计算机自动地从原始文档中提取文摘的技术。自动文本摘要技术的分类方法众多,根据其实现方式的不同可以分为摘录式文摘和理解式文摘^[1]。摘录式文摘通过从原始文档中抽选一些相关的句子并将其重新组合从而形成文摘,这是一种简单有效的自动文本摘要技术;与摘录式文摘不同,理解式文摘生成的文摘中的词语不需要与原始文档相同,因此理解式文摘是使用不同形式对原始文档的中心内容和概念的重新表示。

近几年,递归神经网络(Recurrent Neural Networks, RNN)在自然语言处理任务中展现出优异的性能,例如机器

翻译^[2]、语音识别^[3]和文本压缩^[4]等。这些网络模型避免了复杂的文本分析工作,但模型的训练需要大量的数据。本文主要研究句子级别的文本摘要任务,即标题生成任务。文本自动摘要研究带来了两个方面的挑战:1)文本的主题容易出现漂移或交叉现象,使用常规方法生成的摘要难以全面覆盖事件或话题的内容和主题;2)文本内容本身的社会关系比较复杂,只有深入挖掘文本内容的语义关系才能增强摘要的逻辑性。近些年,人工摘要领域的研究表明,文本摘要是多种文本操作共同作用的结果,例如文本压缩、文本解析、文本生成等。最近,Rush等人^[5]提出了一种单纯由数据驱动的理解式文摘模型,其利用编码-解码结构,使用一个卷积神经网络(Convolutional Neural Network, CNN)作为编码器,并使用一个前馈网络构成的神经网络语言模型(Neural Network Language Model, NNLM)^[6]作为解码器;该模型在UDC2004中取得了优异的成绩。Hu等人^[7]将这种模型结构应用到了中

到稿日期:2017-05-08 返修日期:2017-09-28

庞超 硕士生,主要研究方向为文本摘要与机器翻译;尹传环 副教授,主要研究方向为网络安全(入侵检测)、数据挖掘、机器学习(支持向量机),E-mail:chyin@bjtu.edu.cn(通信作者)。

文文本中,并构建了一个用于训练模型的中文语料库 LCSTS,为研究中文理解式文摘提供了便利。

在分析新闻标题生成方式的过程中,不同类别的新闻标题的生成方式各有特点,例如,对于时政类的新闻标题,标题内容往往倾向于陈述事实;而对于娱乐类的新闻标题,标题内容多使用夸张和歧谬等手法使得标题能够吸引读者的眼球。为了捕获不同类别的新闻在标题生成中的差异特性,我们在标题生成的过程中同时进行文本分类任务,提出了一种基于分类的文本摘要模型。该模型可以在训练文本摘要模型的同时训练文本分类模型,在两部分的共同作用下进一步提高整个模型生成标题的质量。

1 基于分类的文本摘要模型

递归神经网络在语音识别、机器翻译和自动问答系统等自然语言处理任务中显示出了优异的性能。近几年,利用深度学习技术进行自动文本摘要的研究^[8-10]已成为自动文本摘要的热点。本节将详细介绍基于递归神经网络的编码-解码结构,并介绍如何将编码-解码结构与分类结构相结合。

1.1 递归神经网络

递归神经网络是一类用于处理序列数据 $\mathbf{x}=[x_1, x_2, \dots, x_T]$ 的特殊神经网络,由隐藏状态 \mathbf{h} 和可以选择使用的输出 o 构成。在每一个时间步,递归神经网络的隐藏状态 h_t 通过式(1)进行更新,如图 1 所示。

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

其中, f 为非线性激活函数。函数 f 可以选择简单的 sigmoid 函数,也可以选择复杂的长短期记忆(Long Short-Term Memory, LSTM)^[11] 处理单元。

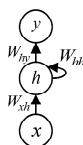


图 1 递归神经网络

Fig. 1 Recurrent neural network

递归神经网络通过训练模型根据当前字符预测序列中的下一字符,使得模型学习到整个序列上的概率分布。在一般情况下,每个时间步的输出是条件概率 $p(x_t | x_{t-1}, \dots, x_1)$ 。例如,当数据中词语的概率分布遵循多项分布时,可以使用 softmax 激活函数表示输出,即对于所有可能的字符 $j = 1, \dots, K$,有:

$$p(x_{t,j} | x_{t-1}, \dots, x_1) = \frac{\exp(w_j h_t)}{\sum_{j=1}^K \exp(w_j h_t)} \tag{2}$$

其中, w_j 是隐藏状态到输出的权重矩阵 \mathbf{W}_{hy} 的第 j 行。最后,通过将 these 概率结合起来计算整个序列的概率,使用式(3)计算序列 \mathbf{x} 的概率:

$$p(\mathbf{x}) = \prod_{t=1}^T p(x_t | x_{t-1}, \dots, x_1) \tag{3}$$

1.2 基于递归神经网络的编码-解码结构

基于递归神经网络的编码-解码结构^[12-14]由读取输入序

列的递归神经网络(称为编码器)以及生成输出序列的递归神经网络(称为解码器)组成。编码器的最终隐藏状态用于计算固定大小的上下文变量 \mathbf{c} 。 \mathbf{c} 表示输入序列的语义概要并作为解码器的输入,解码器将上下文变量重新映射为输出序列,如图 2 所示。为了便于在下文叙述模型的结构,在基于递归神经网络的编码-解码结构中,编码器的隐藏状态使用 \mathbf{h} 表示,解码器的隐藏状态使用 \mathbf{s} 表示。

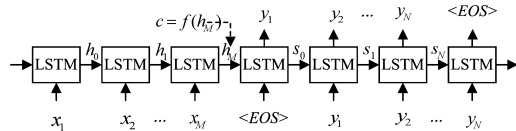


图 2 基于递归神经网络的编码-解码结构

Fig. 2 Encoder-decoder based on recurrent neural network

假设输入序列 $\mathbf{x}=[x_1, x_2, \dots, x_M]$, 输出序列 $\mathbf{y}=[y_1, y_2, \dots, y_N]$, 在文本摘要问题中,一般有 $M > N$ 。从概率的角度上讲,文本摘要问题可以看作在给定输入序列 \mathbf{x} 时,寻找输出序列 \mathbf{y} , 使得输出序列 \mathbf{y} 的条件概率 $p(\mathbf{y} | \mathbf{x})$ 最大。文本摘要问题可以使用式(4)表示。

$$\arg \max_{\mathbf{y} \in Y} \text{score}(\mathbf{x}, \mathbf{y}) \tag{4}$$

其中,集合 Y 表示所有长度为 N 的序列构成的集合;函数 score 表示在给定输入序列 \mathbf{x} 时生成输出序列 \mathbf{y} 的评价函数。文本摘要问题的实质是在集合 Y 中寻找最优输出序列,使得评价函数 score 取得最大值。在实际中,评价函数 score 可以使用给定输入序列 \mathbf{x} 生成输出序列 \mathbf{y} 的对数条件概率,即 $\text{score}(\mathbf{x}, \mathbf{y}) = \log p(\mathbf{y} | \mathbf{x}; \theta)$ 。其中对数条件概率可以表示为:

$$\log p(\mathbf{y} | \mathbf{x}, \theta) = \sum_{i=0}^{N-1} \log p(y_{i+1} | \mathbf{x}, \mathbf{y}_w; \theta) \tag{5}$$

在生成输出序列的过程中使用了 ω 阶马尔科夫假设,即在生成字符 y_{i+1} 时,考虑其之前长度为 ω 的上下文序列 \mathbf{y}_w 。

在编码器中,模型使用递归神经网络按序读取输入序列 \mathbf{x} 中的每一个字符。在读取每个字符的同时,构成编码器的递归神经网络使用式(1)更新隐藏状态,直到读取到特殊字符时(使用特殊字符 $\langle EOS \rangle$ 表示序列的结束),模型利用编码器的最终隐藏状态生成表示输入序列的语义概要的上下文变量 \mathbf{c} 。

在解码器中,模型将编码器输出的上下文变量 \mathbf{c} 作为构成解码器的递归神经网络的初始隐藏状态或输入,通过训练解码器预测其在给定隐藏状态 s_t 生成字符 y_t 的方式使网络按序生成输出序列。在解码器中,隐藏状态使用式(6)进行更新。

$$s_t = f(s_{t-1}, \mathbf{y}_w, \mathbf{c}) \tag{6}$$

其中, \mathbf{y}_w 表示输出序列字符 y_t 的长度为 ω 的上下文字符序列。因此,在解码器上生成下一字符的条件概率为:

$$p(y_t | y_{t-1}, y_{t-2}, \dots, y_1) = g(s_t, \mathbf{y}_w, \mathbf{c}) \tag{7}$$

其中,函数 g 可选择 softmax 等函数进行概率计算,其输入是 $s_t, \mathbf{y}_w, \mathbf{c}$ 的线性组合。

1.3 注意机制

上节中提及的基于递归神经网络的编码-解码结构的一个明显缺陷是编码器输出的上下文变量的维度太小,难以有

效地概括输入序列所包含的信息。为解决这一问题,研究者们引入注意机制(Attention Mechanism)^[15-16],其本质为加权平均,将上下文变量 c 拓展为上下文变量序列 $cs=[c_1, c_2, \dots, c_N]$ 。注意机制使用权重 α_i 的特征向量序列 n 的加权平均表示输出序列字符 y_i 所对应的上下文变量 c_i 。

上下文向量 c_i 依赖特征向量序列 n ,特征向量序列 n 反映了输入序列所蕴含的信息,特征向量序列 n 中的每一项与输入序列的每一项逐一对应,分别表示输入序列中第 i 个词的信息。例如,特征向量序列可以使用编码器隐藏状态序列 $h=[h_1, h_2, \dots, h_M]$ 。

上下文向量 c_i 可以通过对特征向量序列 n 的加权平均来计算:

$$c_i = \sum_{j=1}^N \alpha_{ij} n_j \quad (8)$$

每个特征向量 h_j 的权重 α_{ij} 可通过式(9)来计算:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^N \exp(e_{ik})} \quad (9)$$

其中:

$$e_{ij} = a(s_{i-1}, h_j) \quad (10)$$

式(10)是一个对齐模型(alignment model),表示输入序列中第 j 个字符和输出序列中第 i 个字符的匹配情况。

1.4 基于分类的模型

基于递归神经网络的编码-解码结构的最终隐藏状态是该结构的冗余信息,即在原有结构中没有得到有效利用而最终的隐藏状态携带解码时的特征信息,基于分类的模型充分利用了该冗余信息,将最终的隐藏状态输入分类器中,对输入序列进行分类,并通过分类结果进一步优化文本摘要性能。基于分类的模型如图3所示。

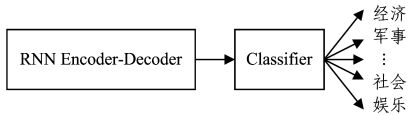


图3 基于分类的文本摘要模型

Fig. 3 Text summarization model based on classification

模型将基于递归神经网络的编码-解码结构的最终隐藏状态输入到分类器中。为简化模型,本文使用 softmax 分类器,如式(11)所示。

$$p(label_i | s_N) = \frac{\exp(\omega_{ij} s_{N,j})}{\sum_{j=1}^K \exp(\omega_{ij} s_{N,j})} \quad (11)$$

其中, ω_{ij} 表示隐藏状态 s_N 的第 j 项与分类器第 i 个输出间的连接权重。

1.5 模型的训练

基于分类的模型由基于递归神经网络的编码-解码结构和分类器组成,这两部分可以在整个训练集上同时训练。我们使用最小批次随机梯度下降(mini-batch stochastic gradient descent)优化方式训练整个模型。整个模型的损失函数如式(12)所示:

$$loss = loss_{Enc-Dec} + \beta loss_{Cla} \quad (12)$$

其中, β 为平衡两部分损失的模型参数, $loss_{Enc-Dec}$ 和 $\beta loss_{Cla}$ 可

分别使用式(13)和式(14)表示。

$$\begin{aligned} loss_{Enc-Dec} &= - \sum_{j=1}^Z \log p(y^{(j)} | x^{(j)}; \theta) \\ &= - \sum_{j=1}^Z \sum_{i=0}^N \log p(y_{i+1}^{(j)} | x^{(j)}, y_w^{(j)}; \theta) \end{aligned} \quad (13)$$

$$loss_{Cla} = - \sum_{j=1}^Z \log p(label^{(j)} | s_N^{(j)}; \epsilon) \quad (14)$$

在式(13)和式(14)中,模型使用的数据集可以表示为 $\{\langle x^{(j)}, y^{(j)}, label^{(j)} | j=1, \dots, Z \rangle\}$ 。其中 $x^{(j)}$ 表示输入序列, $y^{(j)}$ 表示输出序列, $label^{(j)}$ 表示文本分类的标签。

2 实验

基于分类的模型完成了两种自然语言处理任务:1)文本摘要任务;2)文本分类任务。因此我们实现了两组对比实验,在一组实验中对比了基于分类的模型和非分类的模型在文本摘要任务中的性能;在另一组实验中对比了基于分类的模型和基于朴素贝叶斯的分类算法在文本分类任务中的性能。另外,本节详细描述了在实验中使用的数据集和评价方法。

2.1 数据集

通过爬虫技术从中国新闻网获取新闻内容,形成语料库。该语料库共有 120 万条语料,其中训练集包括 90 万条语料,验证集包括 20 万条语料,测试集包括 10 万条语料。每条语料由 3 部分组成:新闻的标题、新闻的内容和新闻的类别(共分为 11 个类别,包括时政、国际、社会、财经、金融、汽车、能源、文化、娱乐、体育、健康)。对话料库中的文本内容进行预处理:

(1)去除文本中的特殊字符;

(2)将文本中的数字替换为 NUM_标签,将日期替换为 DATA_标签;

(3)使用两种方式对文本进行分词:1)使用 jieba 对文本进行分词;2)按照字符对文本进行分割。

2.2 评价标准

本文提出的模型实际上完成了两种自然语言处理任务:1)文本摘要任务;2)文本分类任务。本节介绍文本摘要和文本分类的评价方法。

(1)文档摘要评价方法

文档摘要评价方法大致分为两类:1)内部评价方法(Intrinsic Methods)。该方法提供参考摘要,以参考摘要为基准评价系统生成摘要的质量。系统生成摘要与参考摘要越吻合,质量越高。2)外部评价方法(Extrinsic Methods)。该方法不提供参考摘要,利用文档摘要代替原始文档执行某个文档相关的应用,例如文档检索、文档聚类、文档分类等,能够提高应用性能的摘要被认为是高质量的摘要。本文使用一种常用于自动文本摘要的相关国际评测的内部评价方法 ROUGE^[17]。

ROUGE 是由 ISI(Information Sciences Institute)的 Lin 和 Hovy 提出的一种自动文本摘要评价方法,现被广泛应用于 DUC(Document Understanding Conference)的摘要评测任务中。

ROUGE 基于摘要中 n 元词(n -gram)的共现信息来评价

摘要质量,是一种基于 n 元词召回率的评价方法。ROUGE 由一系列的评价方法组成,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:

$$\frac{\sum_{S \in \{RefSummaries\}} \sum_{n\text{-gram} \in S} Count_{match}(n\text{-gram})}{\sum_{S \in \{RefSummaries\}} \sum_{n\text{-gram} \in S} Count(n\text{-gram})} \quad (15)$$

其中, $n\text{-gram}$ 表示 n 元词, $\{RefSummaries\}$ 表示参考摘要, $Count_{match}(n\text{-gram})$ 表示系统生成摘要和参考摘要中同时出现 $n\text{-gram}$ 的个数, $Count(n\text{-gram})$ 则表示参考摘要中出现 $n\text{-gram}$ 的个数。ROUGE-L 表示系统生成摘要和参考摘要的最长公共序列的长度与参考摘要的长度的比值。相比于 ROUGE-N, ROUGE-L 考虑了摘要中词语的次序,评价更为合理。

(2) 文本分类评价方法

对于文本分类任务,本文使用精确率 (precision) 和召回率 (recall) 作为文本分类性能的评价指标。

$$precision = \frac{TP}{TP + FP} \quad (16)$$

$$recall = \frac{TP}{TP + FN} \quad (17)$$

其中, TP 表示真正例, FP 表示假正例, FN 表示假负例。

2.3 实现

为训练模型,使用最小批次随机梯度下降来最小化负对数概率(见式(12)),设置学习率为 0.5,批次大小为 64,模型隐藏层单元数量为 1024,式(12)中的参数 α 取值为 0.5。为了更好地捕获词语的语义信息,模型的输入使用了词语的词向量(word vector)。分别在经过词语级分词和经过字符级分词处理的语料上进行实验。对于词语级的文本摘要模型,设置词典大小为 20000;对于字符级的生成模型,设置词典的大小为 5000。模型使用谷歌第二代开源人工智能系统 TensorFlow^[18]实现。在所有的实验中,构成模型的递归神经网络均使用 LSTM 单元,模型的所有参数采用随机初始化。

2.4 实验结果

所有模型均在 NVIDIA GTX TITAN X 上训练,表 1 列出了实验结果,其中 Enc-Dec 表示使用了基于递归神经网络的编码-解码结构,Enc-Dec+Cla 表示使用了由基于递归神经网络的编码-解码结构和分类器组成的基于分类的文本摘要模型,Word 表示在经过词语级分词处理的语料上进行实验,Char 表示在经过字符级分词处理的语料上进行实验。从实验结果中可以看到,在词语级和字符级实验中,基于分类的模型的性能优于基于递归神经网络的编码-解码结构的模型。该结果表明通过增加分类结构的模型可以进一步优化模型生成摘要的性能,使得模型按照类别捕获到不同类别文本的摘要特点。另外,从实验结果中可以看出,字符级的模型结构优于词语级的模型结构。在词语级的模型结构中,模型生成的 UNK_ 多于字符级的模型结构的结果。字符级的模型减少了 UNK_ 的出现次数,一定程度上解决了低频词问题,同时也避免了在预处理阶段一些错误的分词对摘要模型所带来的影响,提高了系统的鲁棒性。

表 1 文本摘要实验结果

名称	ROUGE-1	ROUGE-2	ROUGE-L
Enc-Dec(Word)	25.01	7.34	12.50
Enc-Dec(Char)	28.72	8.41	15.43
Enc-Dec+Cla(Word)	27.28	8.36	13.89
Enc-Dec+Cla(Char)	31.44	11.15	16.67

在文本分类的对比实验中(见表 2),所提出的基于分类的文本摘要模型的性能和基于朴素贝叶斯的文本分类模型的性能近似,其中词语级模型的性能高于字符级模型。

表 2 文本分类实验结果/%

名称	精确率	召回率
Naive Bayesian	70.34	83.98
Enc-Dec+Cla(Word)	69.97	81.63
Enc-Dec+Cla(Char)	67.86	82.99

在模型的编码-解码结构中使用的对齐模型的可视化结果如图 4 所示。在图 4 的实例中,每一行表示原文中的词语对应生成摘要中各词语的权重值,颜色越深,权重值越大。可以看出,在原文和摘要中相关的词有着更高的值。

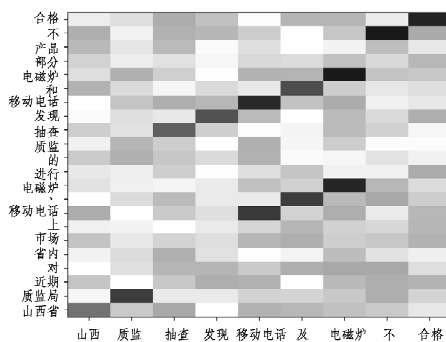


图 4 对齐模型实例

Fig. 4 A example of alignment

将基于分类的文本摘要模型和几种优秀的文摘模型进行对比,所有模型均在经过词语级分词的语料上进行训练和测试,结果如表 3 所列。

表 3 对比实验结果

名称	ROUGE-1	ROUGE-2	ROUGE-L
MOSES+	24.50	8.13	12.85
RAS-Elman	26.97	8.46	13.06
RNN-context	25.03	7.52	12.41
Enc-Dec+Cla	27.28	8.36	13.89

MOSES+^[5]:一种基于 MOSES 的标题生成系统,其中 MOSES^[19]是一种基于解析的机器翻译模型。

RAS-Elman^[20]:利用一个卷积神经网络作为编码器,编码器以词语和词语位置作为输入,并且使用基于注意机制的解码器,使用 Elman 递归神经网络^[21]作为解码器。

RNN-context^[7]:一种简单的基于编码-解码结构的文摘模型,该模型将在编码器中的所有隐藏状态链接起来作为编码器的输入。

孙林,刘弱南,张霄雨,等.一种基于粗糙均方残基的模糊双聚类方法[J].河南师范大学学报(自然科学版),2017,45(5):93-100.

- [15] HU Q H, ZHAO H, YU D R. Efficient symbolic and numerical attribute reduction with neighborhood rough sets[J]. Pattern Recognition & Artificial Intelligence, 2008, 21(6): 732-738. (in Chinese)

胡清华,赵辉,于达仁.基于邻域粗糙集的符号与数值属性快速约简算法[J].模式识别与人工智能,2008,21(6):732-738.

- [16] HU Q H, YU D R, LIU J F, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2008, 178(18): 3577-3594.

- [17] XUE Z A, WANG N, SI X M, et al. Research on multi-granularity rough intuitionistic fuzzy cut sets[J]. Journal of Henan Normal University (Natural Science Edition), 2016, 44(5): 131-139. (in Chinese)

薛占熬,王楠,司小滕,等.多粒度粗糙直觉模糊截集的研究[J].河南师范大学学报(自然科学版),2016,44(5):131-139.

(上接第147页)

所有实验均在经过词语级分词处理的语料上进行,基于分类的模型得到了较好的结果。

结束语 本文构建了基于分类的文本摘要模型,该模型将基于递归神经网络的编码-解码机构与分类器相结合,并在大量的语料下同时训练优这两部分,从而在文本摘要任务中取得了优异的性能。但在模型中低频词的问题仍然存在,特别是在词语级的实验中,本文通过使用字符级的分词处理来解决该问题。近年来,研究者们提出了神经图灵机(Neural Turing Machines, NTM)^[22]的概念,其已在许多问题中表现出了极佳的性能,在今后的工作中将尝试把神经图灵机融入到文本摘要模型中,以获得更好的文本摘要性能。

参考文献

- [1] GAMBHIR M, GUPTA V. Recent automatic text summarization techniques: a survey [J]. Artificial Intelligence Review, 2017, 47(1): 1-66.
- [2] 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. Lisbon: ACL, 2015: 1412-1421.
- [3] GRAVES A, MOHAMMED A R, HINTON G. Speech recognition with deep recurrent neural networks IEEE[C]// International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 6645-6649.
- [4] GAMBHIR M, GUPTA V. Recent automatic text summarization techniques: a survey [J]. Artificial Intelligence Review, 2017, 47(1): 1-66.
- [5] RUSH A M, CHOPRA S, WESTON J. A Neural Attention Model for Abstractive Sentence Summarization [C]// Proceedings of NAACL. 2016.
- [6] BENGIO Y, SCHWENK H, SENÉCAL J, et al. Neural Probabilistic Language Models[J]. Journal of Machine Learning Research, 2006, 3(6): 1137-1155.
- [7] HU B, CHEN Q, ZHU F. LCSTS: A Large Scale Chinese Short Text Summarization Dataset [C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015: 1967-1972.
- [8] LOPYREV K. Generating News Headlines with Recurrent Neural Networks[J]. Computer Science, 2015.
- [9] CHOPRA S, AULI M, RUSH A M. Abstractive Sentence

Summarization with Attentive Recurrent [C]// Neural Networks Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. 2016: 93-98.

- [10] NALLAPATI R, ZHOU B, SANTOS C N D, et al. Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond[J]. CoNLL, 2016, 1(1): 280-290.
- [11] SURHONE L M, TENNOE M T, HENSSONOW S F. Long Short Term Memory [C]// Betascript Publishing. 2010.
- [12] CHO K, VAN M B, GULCEHRE C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014, 1(1): 43-66.
- [13] SUTSKEVER I, VINYALS O, LE Q V. Sequence to Sequence Learning with Neural Networks[J]. Advances in Neural Information Processing Systems, 2014, 4: 3104-3112.
- [14] BENGIO S, VINYALS O, JAITLEY N, et al. Scheduled sampling for sequence prediction with recurrent Neural networks[C]// International Conference on Neural Information Processing Systems. MIT Press, 2015: 1171-1179.
- [15] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[C]// Proceedings of Neural Information Processing Systems 2014. NIPS, 2014.
- [16] VINYALS O, KAISER L, KOO T, et al. Grammar as a foreign language[J]. Eprint Arxiv, 2014, 1(1): 2773-2781.
- [17] 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. Association for Computational Linguistics, 2003: 71-78.
- [18] Google. Tensorflow (Version 1. 2) [OL]. <http://www.tensorflow.org>.
- [19] KOEHN P, HOANG H, BIRCH A, et al. Moses; Open source toolkit for statistical machine translation[C]// Proceedings of ACL. 2007: 177-180.
- [20] CHOPRA S, AULI M, RUSH A M, et al. Abstractive sentence summarization with attentive recurrent neural networks[C]// Proceedings of NAACL. 2016.
- [21] ELMAN J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2): 179-211.
- [22] GRAVES A, WAYNE G, DANIHELKA I. Neural Turing Machines[J]. Computer Science, 2014, 1(1): 89-95.