计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.181202261

一种基于注意力机制的中文短文本关键词提取模型

杨丹浩　吴岳辛　范春晓
北京邮电大学电子工程学院　北京 100089
(danhaoy94@163.com)

摘　要　关键词抽取技术是自然语言处理领域的一个研究热点。在目前的关键词抽取算法中,深度学习方法较少考虑到中文的特点,汉字粒度的信息利用不充分,中文短文本关键词的提取效果仍有较大的提升空间。为了改进短文本的关键词提取效果,针对论文摘要关键词自动抽取任务,提出了一种将双向长短时记忆神经网络(Bidirectional Long Shot-Term Memory,BiLSTM)与注意力机制(Attention)相结合的基于序列标注(Sequence Tagging)的关键词提取模型(Bidirectional Long Short-term Memory and Attention Mechanism Based on Sequence Tagging,BAST)。首先使用基于词语粒度的词向量和基于字粒度的字向量分别表示输入文本信息;然后,训练 BAST 模型,利用 BiLSTM 和注意力机制提取文本特征,并对每个单词的标签进行分类预测;最后使用字向量模型校正词向量模型的关键词抽取结果。实验结果表明,在 8159 条论文摘要数据上,BAST 模型的 *F*1 值达到 66.93%,比 BiLSTM-CRF(Bidirectional Long Shoft-Term Memory and Conditional Random Field)算法提升了 2.08%,较其他传统关键词抽取算法也有进一步的提高。该模型的创新之处在于结合了字向量和词向量模型的抽取结果,充分利用了中文文本信息的特征,可以有效提取短文本的关键词,提取效果得到了进一步的改进。

关键词:注意力机制;词向量;字向量;关键词抽取;LSTM
中图法分类号　TP391

Chinese Short Text Keyphrase Extraction Model Based on Attention

YANG Dan-hao,WU Yue-xin and FAN Chun-xiao
School of Electronic Engineering,Beijing University of Posts and Telecommunications,Beijing 100089,China

Abstract　Keyphrase extraction technology is a research hotspot in the field of natural language processing. In the current keyphrase extraction algorithm,the deep learning method seldom takes into account the characteristics of Chinese,the information of Chinese character granularity is not fully utilized,and the extraction effect of Chinese short text keyworks still has a large improvement space. In order to improve the effect of the keyphrase extraction for short text,a model for automatic keyphrase extraction abstracts was proposed,namely BAST model,which combines the bidirectional long short-term memory and attention mechanism based on sequence tagging model. Firstly, word vectors in the word granularity and character vectors in the character granularity are used to represent input text information. Secondly,the BAST model is trained,text features are extracted by using BiLSTM and attention mechanism,and the label of each word is classified. Finally,the character vector model is used to correct the extraction results of the word vector model. The experimental results show that the F1-measure of the BAST model reaches 66.93% on 8159 abstract data,which is 2.08% higher than that of the BiLSTM-CRF(Bidirectional Long Shoft-Term Memory and Conditional Random Field) algorithm,and is further improved than other traditional keyphrase extraction algorithms. The innovation of the model lies in the combination of the extraction results of the word vector and the character vector model. The model makes full use of the characteristics of the Chinese text information and can effectively extract keyphrases from the short text,and extraction effect is further improved.

Keywords　Attention mechanism,Word embedding,Character embedding,Keyphrase extraction,LSTM

1　引言

关键词提取技术是一种可以从文本中抽取主题和一些重要短语的技术,可以帮助阅读者快速了解文本中最有价值的信息。

论文的摘要作为文本的精华部分,可以简要地概括整篇文本的主要内容,且篇幅较短。论文文本的关键词选取比较严谨,方便作为数据集使用,因此经常被用于关键词提取任务。Gollapalli 等[1]利用图算法从英文论文的题目和摘要中提取关键词。2017 年,Florescu 等[2]利用无监督学习方法从学术论文的标题和摘要数据集上提取关键词。

从论文中高效快速地提取关键词,可以方便研究人员根

到稿日期:2018-12-05　返修日期:2019-04-25　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:范春晓(fcxg100@163.com)

据关键词查找论文文献，掌握该领域内的最新研究成果，同时也有助于对文献做分类聚类，方便数据的管理和使用。由此，本文面对中文论文文献，针对论文摘要关键词自动抽取任务，基于深度学习算法建立关键词自动提取模型。

关键词提取主要使用有监督和无监督两种算法，其中机器学习、神经网络等算法取得了较好的效果。Hasan 等[3]利用 TF-IDF 的改进算法，结合文本的多项特征来提取关键词。基于图的无监督算法计算词与词的共现关系，并且可以结合文档的其他特征为每个单词打分，根据打分高低从文本中抽取关键词。近年来，研究者提出了许多基于图的算法，并引入了文本的各种特征，对算法作出了改进，如 TextRank[4]，TopicRank[5]，SalienceRank[6]，PositionRank[7] 等算法。Zhang 等[8]使用条件随机场(Conditional Random Field，CRF)，根据词频、位置、词性标注等文本特征提取关键词，并与支持向量机等模型做对比。Haddoud 等[9]使用逻辑回归分类器来提取关键词，用二分类的方法判断单词类别。Onan 等[10]利用集成学习的方法，结合多种有效的机器学习算法来提取关键词，并将提取结果应用于文本分类问题。2017 年，Gollapalli 等[11]结合多种机器学习算法，利用这些信息进一步改进了条件随机场算法的关键词提取效果。

随着深度学习研究的深入，研究者开始使用神经网络提取关键词。2015 年，Zhang 等[12]提出使用一个双层的循环神经网络(RNN)来提取 Twitter 短文本的关键词，其中每条推文有唯一的关键词。LSTM-CRF 模型[13]可用于解决序列标注问题，从而更好地提取上下文特征。在此基础上，Mourad[14]利用基于 LSTM-CRF 的单词和字母级别的文本信息，在生物医学命名实体识别的多个英文语料库中取得了不错的效果。2017 年，Andrej 等进一步提出基于 self-attention 机制[15]的命名实体识别模型，该模型利用一个多头的编码解码器并加入注意力机制，来实现推特文本主题提取任务。

综上所述，现有研究较少涉及中文短文本的关键信息提取。鉴于此，本文研究了文本特征表示的方法，结合双向长短时记忆神经网络以及注意力机制，建立了 BAST 神经网络模型。该模型结合了中文的词向量和字向量预测结果，同时利用单词级和字级的文本信息来提高模型的关键词抽取能力。实验在大量的中文论文摘要数据集上进行，并对比了多种已有模型。本文模型在多关键词的论文摘要关键词自动抽取任务中的准确率达到 66.39%，召回率达到 67.48%，$F1$ 值达到 66.93%，具有较好的提取效果。

2 BAST 关键词抽取模型

2.1 BAST 模型结构

本文设计的 BAST 关键词提取模型利用 BiLSTM 模型处理长距离依赖问题，并结合注意力机制，进一步提取了上下文信息，最后利用两种不同粒度的文本表示方法，共同抽取文本关键词。模型整体结构如图 1 所示，该模型包括文本词向量输入层、文本字向量输入层、BiLSTM 神经网络层、注意力机制层、标签分类层和预测结果过滤层。

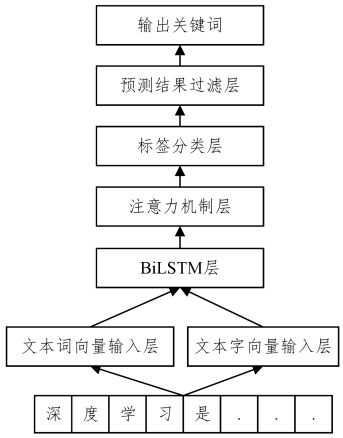

图 1 BAST 模型的结构框图

Fig. 1 Structure diagram of BAST model

2.2 文本词向量输入层

输入的原始文本的形式为汉字，无法被神经网络直接训练，因此需要转化为向量形式。文本词向量输入层是模型的最底层，输入为需要提取的关键词的文本，输出为由文本序列转化成的词向量序列，用于神经网络训练。

一种向量化的方法是 one-hot 方法，在特征提取上属于词袋模型，不考虑词与词之间的顺序，且维度过大，无法体现单词间的语义相关性。Si 等[16]使用将单词转化为词向量的方法，这是一种可以弥补 one-hot 方法的缺点的新型单词表示方法。使用 Word2Vec 模型，可以将文本中的字词转化为一定维度的向量，以方便神经网络进行高效的特征提取计算。Word2Vec 模型包括 Skip-gram 和 CBOW 两种训练方式，CBOW 模型利用上下文单词预测中心单词 W_t，而 Skip-gram 是在已知单词 W_t 的情况下预测其上下文单词 W_{ct}。本文选用 Skip-gram 方式来训练词向量。

2.3 文本字向量输入层

为了弥补词向量模型在训练时的不足，本文使用基于汉字粒度的向量进一步提取文本特征。字向量是指将中文文本按照每个字来训练，为每个汉字生成一个具有一定维度的向量。字向量可以更好地表达每个汉字的含义，同样可以作为神经网络的输入，成为词向量的一种有效补充。

字向量模型的输出可作为预测结果过滤层的输入，用于筛选词向量模型输出的候选关键词，共同产生最终的关键词预测结果。

2.4 BiLSTM 神经网络层

本层接收文本词向量层的词向量输出，并采用合适的神经网络模型来提取文本特征。经过实验，本文采用双向长短期记忆网络(BiLSTM)结构，这是一种对长短期记忆网络(Long Short-Term Memory，LSTM)的改进结构。

LSTM[17]是一种改进的循环神经网络，可以较好地解决 RNN 的长期依赖问题。LSTM 网络由 3 个门结构(输入门、遗忘门、输出门)和 1 个状态单元组成。输入门接收两个输入，即上一时刻 LSTM 的输出结果 h_{t-1} 和当前时刻的输入 x_t。t 时刻的输入门的输出计算公式为：

$$i_t = f(\boldsymbol{W}_i x_t + \boldsymbol{W}_i h_{t-1} + b_i) \tag{1}$$

其中，$\boldsymbol{W}$ 表示权重矩阵，b 为偏置向量。

遗忘门同样接收上述两个输入，并决定是否从状态单元中丢弃信息，输出计算式为：

$$f_t=\sigma(\boldsymbol{W}_f x_t+\boldsymbol{W}_f h_{t-1}+b_f) \tag{2}$$

当前时刻的状态单元接收输入门和遗忘门的值，可以表示为：

$$C_t=f_t C_{t-1}+i_t \tilde{C}_t \tag{3}$$

其中，$\tilde{C}_t$ 表示前一时刻的状态单元值，表示为：

$$\tilde{C}_t=\tanh(\boldsymbol{W}_C h_{t-1}+\boldsymbol{W}_C x_t+b_C) \tag{4}$$

输出门用于控制 LSTM 状态单元的输出，表达式为：

$$o_t=f(W_o x_t+\boldsymbol{W}_o h_{t-1}+b_o) \tag{5}$$

当前单元的输出表示为：

$$h_t=o_t\tanh(C_t) \tag{6}$$

BiLSTM[18]是对 LSTM 神经网络的改进，能更有效地利用文本序列的上下文信息，可以更好地提取文本特征。本文利用 BiLSTM 神经网络结构来整合两个方向的 LSTM 神经网络的输出，并将其拼接起来作为整体传入下一层。

将 t 时刻的前向神经网络的输出表示为 h_{ti}，反向神经网络输出表示为 h_{tj}，则在 t 时刻，BiLSTM 的最终输出可以表示为：

$$h_t=\{h_{ti},h_{tj}\} \tag{7}$$

2.5 注意力机制层

本层将上一层 BiLSTM 的输出作为输入，利用注意力机制进一步提取文本特征。在自然语言处理领域，注意力机制[19]最早应用于机器翻译任务中，是一种可以提取当前节点更重要信息的机制。对于短文本，词汇量一般不超过 100，每个单词对整个文本关键词的重要程度是不一样的，注意力机制可以重新分配单词权重，以进一步提取文本中更深层次的信息。最新研究表明[20]，注意力机制在序列标注问题上也可以取得较好的效果，因此可以将注意力机制与 BiLSTM 神经网络相结合，以进一步提取特征，有效地突出文本关键词权重。

首先定义一个注意力机制的矩阵，将 BiLSTM 神经网络的输出作为输入，通过非线性变换得到节点 i 对节点 j 的隐含表示 e_{ij}：

$$e_{ij}=\boldsymbol{V}\tanh(\boldsymbol{W}h_i+\boldsymbol{U}h_j+b) \tag{8}$$

其中，h_i 和 h_j 分别表示前向和反向 LSTM 神经网络的输出，$\boldsymbol{V},\boldsymbol{W},\boldsymbol{U}$ 是权重矩阵。在 n 个时间节点中，第 i 个节点对第 j 个节点的注意力概率权重可以表示为：

$$a_{ij}=\frac{\exp(e_{ij})}{\sum_{k=1}^{n}\exp(e_{ik})} \tag{9}$$

根据权重，可以计算得到第 i 个单词的新的输出特征值：

$$h_{ai}=\sum_{k=1}^{n}a_{ij}h_i \tag{10}$$

采用这种方法同样可以计算第 i 个单词的新的反向 LSTM 特征值 h_{aj}，因此第 t 个单词在进入注意力机制后对应的输出特征为：

$$h_t=\{h_{ai},h_{aj}\} \tag{11}$$

2.6 标签分类层

为了将关键词提取问题转化为分类问题，本文定义一个输出层，为每个单词判断标签。单词的标签为 T 和 F，分别表示当前单词属于关键词和不属于关键词。这一层可以把之前输出的高维度特征映射到低维度的类别上，并经过 softmax 分类器计算对应类别的分布概率，表达式为：

$$y_i=\text{softmax}(\boldsymbol{W}_c H+b_c) \tag{12}$$

其中，H 表示注意力层的输出，$\boldsymbol{W}$ 为权重矩阵，b 是偏置向量。在训练时，目标为最小化损失函数，损失函数为 softmax 的输出向量和样本的正确标签的交叉熵损失：

$$H_{y'}(y)=-\sum_i y_i'\log(y_i) \tag{13}$$

其中，y_i'表示第 i 个正确标签的值，y_i 表示 softmax 的输出向量中的第 i 个标签的值。通过这一输出层得到第 i 个单词对应的预测标签。

2.7 预测结果过滤层

本文在标签分类层的输出的基础上，设计了一个预测结果过滤层。这一层可以结合词向量预测模型和字向量预测模型两种粒度的预测结果，来修正以词向量为主的预测模型，从而提高预测准确率和召回率。

经过反复实验发现，随着训练周期的增加，模型可能会出现一定的过拟合现象，表现为预测关键词数量可能多于实际关键词数量，模型的召回率上升、准确率下降而 $F1$ 值难以进一步提升。为了更好地预测关键词、减轻过拟合的现象，引入字向量模型的训练结果，结合词向量和字向量的预测结果，将二者共同预测的单词作为最终的预测结果。过滤层的模型结构如图 2 所示。

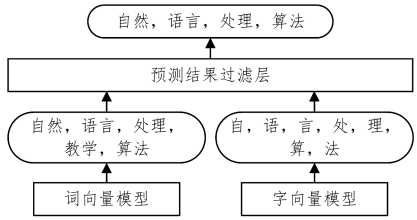

图 2 预测结果过滤层示意图

Fig. 2 Diagram of prediction result filter layer

模型将输入文本转化为词向量粒度的输入和字向量粒度的输入，并分别送入神经网络中进行训练。将词向量作为输入，则模型的标签和模型的预测结果也是单词。同理，将字向量作为输入，则模型的标签为汉字，模型的输出同样是汉字。词向量模型与字向量模型均可以较好地提取关键词。设 O_w 为预测关键词的集合，由 n 个词语组成，可表示为 $O_w=\{y_1,y_2,\cdots,y_n\}$，将其中的每个单词都拆分成汉字，共拆分为 t 个单词，结果为 $O_w=\{c_{w1},c_{w2},\cdots,c_{wt}\}$。$O_c$ 为预测关键字的集合，由 m 个汉字组成，可表示为 $O_c=\{c_1,c_2,\cdots,c_m\}$，则最终输出的 O 可以表示为：

$$O=filter(O_w,O_c) \tag{14}$$

本文定义一个过滤函数，取词向量模型的全部关键词预测结果 O_w 作为初步的预测结果，然后与字向量模型的预测结果 O_c 进行对比，如果词向量预测的单词中的每个字在字向量模型中都没有得到预测，则在最终的预测结果中过滤该单词；只要有一个字在自向量模型中得到预测，则保留整个预测单词。

如图 2 所示，词向量模型的预测结果为{“自然”，“语言”，“处理”，“教学”，“算法”}，字向量模型的预测结果为{“自”，“语”，“言”，“处”，“理”，“算”，“法”}，“自然”这个词因为字向量模型中出现了“自”得以保留，而“教学”这个词因为字向量中没有出现其中的任何一个字，所以被舍弃。模型最终的预测结果为：{“自然”，“语言”，“处理”，“算法”}。

3 实验与结论

3.1 实验数据

从知网收录的大量中文论文中选取 1 万多篇论文的摘要和关键词，筛选其中符合要求的数据作为训练和测试的数据集，具有包括计算机科学、社会科学、工程技术等多个门类的论文，主题范围较广，文本内容丰富。训练样本共有 8159 条，测试样本有 839 条，其他的数据作为训练集。本文直接采用论文文献的关键词作为关键词提取任务的提取目标。

首先进行文本预处理。为了中文文本训练的方便性，去除文本中的英文单词和特殊符号，使用 Python 正则表达式筛选出中文文本及关键词，直接去除不符合中文规范的样本。接着对文本进行分词，使用 jieba 分词工具得到中文分词后的训练样本数据。然后根据关键词为序列做标注。遍历样本中的每一个单词，如果当前样本属于关键词，则对应的标签为 T，否则为 F。

综上所述，文本预处理算法流程如下：

(1)过滤样本中非中文的单词和特殊字符，只保留中文简体字的样本。

(2)去除长度过长的文本，人工检查关键词的设置情况，去除关键词与文本内容不匹配的文本，去除重复文本。

(3)对文本的每个句子做分词，对关键词中的短语做分词。

(4)根据关键词为文本的每个单词做标注。过滤掉关键词与文本内容不匹配的数据，保存标注完成的文本数据集和关键词。

按照比例将经过预训练后保留的文本随机划分为训练集、测试集、验证集，各个集合的样本数量和关键词个数如表 1 所列。模型随机抽取总数据量的 10%作为测试集，从剩余数据中取出 10%作为验证集，其余文本数据为训练集。

表 1 数据集统计表

Table 1 Dataset statistics table

集合	样本总数	关键词总数
训练集	6608	35151
验证集	735	3926
测试集	816	4227

3.2 实验结果

本文采用准确率(P)、召回率(R)和 F1 值(F)作为模型效果的评价指标，计算式分别为：

$$P=|C|/|E| \tag{15}$$

$$R=|C|/|S| \tag{16}$$

$$F=2PR/(P+R) \tag{17}$$

其中，C 为正确提取的关键词集合，E 为提取的关键词集合，S 为标注关键词的集合。

为了更全面地对比算法的提取效果，首先使用 TF-IDF 和 TextRank 等通用的无监督学习关键词提取算法作为对比算法。TF-IDF 统计每个单词在每条数据中的频率和该单词在整个数据集各个文档中的出现次数，从而衡量一个单词对于这篇文章的重要性。TextRank 算法通过设定窗口，统计单词间的共现次数，从而为每一个单词打分，按得分高低选取预测关键词。在使用 TF-IDF 和 TextRank 算法时，根据关键词的数量变化来抽取关键词。例如，如果一篇摘要有 3 个关键词，则选取预测得分最高的前 3 个单词作为最终预测的关键词。

本文不仅对比了当前关键词提取的主要方法，如 TF-IDF 和 TextRank，同时也对比了深度学习领域常用的神经网络结构，如 RNN，LSTM，GRU 等。BiLSTM-CRF[13] 是一种常用的具有良好效果的序列标注模型，双向的神经网络更有助于利用上下文的信息，增加的 CRF 层则可以学习到标签间的顺序关系，可以舍弃不合理的标签顺序，适合做输出层。

BAST 模型的一些参数对提取效果可能会有较大影响，经过反复实验测试，本文设定模型的词向量维度为 120，前向和反向的 LSTM 隐藏层节点数为 180，学习率为 0.01，dropout 值为 0.5，每次训练的 batch 尺寸为 50 时，模型的实验效果达到最佳。

为了更好地对比各种神经网络模型的提取效果，所有方法的词向量均使用相同维度的词向量，所有网络的隐藏层节点数、dropout 值也都取相同值。因此，本文共进行了 9 个实验，除实验 9 外，其他实验均选择目前通用的关键词提取模型与 BAST 模型进行对比。表 2 列出了每一个模型在测试集上的准确率、召回率和 $F1$ 值，图 3 为实验结果折线图。

表 2 各模型的实验结果对比表

Table 2 Comparison of experimental results of each model

(单位：%)

编号	算法名称	准确率	召回率	$F1$ 值
1	TextRank	49.01	49.01	49.01
2	TF-IDF	54.48	54.48	54.48
3	CRF	53.02	59.49	56.07
4	RNN	62.67	57.31	59.87
5	GRU	60.23	63.03	61.60
6	LSTM	62.53	61.83	62.18
7	BiLSTM	60.77	66.32	63.42
8	BiLSTM+CRF	62.90	66.93	64.85
9	BAST	66.39	67.48	66.93

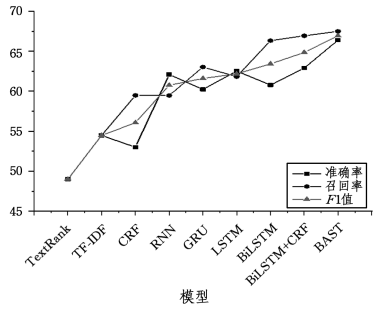

图 3 实验结果对比图

Fig. 3 Comparison of experimental results

实验 1 和实验 2 表明，无监督学习的 TF-IDF 算法和 TextRank 算法取得了相近的提取效果。实验 3 使用有监督

学习的条件随机场(CRF)模型，可以对标签上下文信息进行提取，取得了53.02%的准确率和59.49%的召回率，$F1$值提高到56.07%。

相较于无监督的学习方法，作为有监督模型的神经网络模型在关键词提取问题上取得了更好的效果。实验4中标准的单层RNN神经网络在准确率、召回率、$F1$值上均明显高于CRF模型，F1值达到59.87%。实验5使用的标准的GRU模型和实验6使用的LSTM模型均是对RNN模型结构的改进模型，$F1$值比RNN稍高。GRU为LSTM的一种变体，其关键词提取效果与LSTM相近。实验7中的BiLSTM模型可以从两个方向获得上下文的关系特征，抽取关键词的效果得到进一步提升，$F1$值比实验6的LSTM模型提高了1.24%。

实验结果表明，在神经网络的输出层添加注意力层或CRF层都可以起到提高$F1$值的效果。实验8在BiLSTM神经网络的输出层加上了CRF层，$F1$值达到64.85%。实验9采用本文提的BAST模型。实验结果表明，BAST模型的$F1$值比第二高的BiLSTM-CRF模型的$F1$值提高了2.08%，达到了66.93%，成为关键词提取效果最好的模型。

3.3 对比分析

神经网络的多个参数均可能对关键词的提取效果产生影响，如向量维度、学习率、隐藏层神经元节点数量以及dropout值等。学习率的值不应设置得过大，取0.01时可以得到较好的效果；dropout表示随机遗忘的神经元比例，其取值对模型训练效果的影响不大，取0.5即可达到预期效果。下面对词向量维度和隐藏层的节点数做重点讨论。

首先分析词向量的维数对模型效果的影响，实验结果如图4所示。可以看到，词向量的维数越高，信息量就越大，可以更好地将单词的特征表现出来。当词向量小于100维时，随着词向量维度的上升，模型的$F1$值会提高，但当词向量维数继续增大时，$F1$值维持在0.65～0.67之间，准确率和召回率的变化幅度不大，模型的训练效果基本稳定。可见，词向量在100维以上时，模型更有可能得到较好的训练结果。

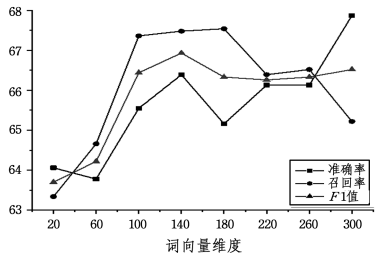

图4 词向量维度与模型训练效果的关系

Fig. 4 Relationship between dimension of word vector and training effect of model

然后分析BiLSTM的隐藏层节点数对关键词提取效果的影响。隐藏层的神经元节点数的变化与关键词预测的准确率、召回率、F1值间的关系如图5所示。当神经元较少时，模型训练速度较快，收敛速度较快，但是提取效果不好。在节点数小于100时，随着神经元数量的增多，$F1$值会不断提升，在节点数超过100后，$F1$值保持在0.65～0.67，模型具有较好的训练效果。

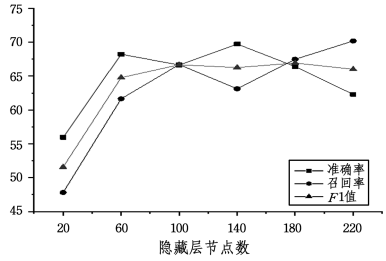

图5 神经元节点数与模型训练效果的关系

Fig. 5 Relationship between number of neuron nodes and training effect of model

接下来讨论使用字向量过滤模型对关键词提取效果的影响。在训练过程中，随着训练轮次的增长，模型出现过拟合现象，$F1$值呈现先上升后下降的趋势。因此，本文设计了只使用词向量的模型和字向量与词向量相结合的模型的对比实验，在其他参数设置均相同的情况下，观察模型在不同的全局训练轮次下的训练效果。WV表示只使用词向量的模型，WV+CV表示增加了字向量过滤层后的模型，两个模型均使用注意力机制，并通过softmax函数输出预测结果，其在不同的全局步数下的$F1$值对比图如图6所示，每一个全局步代表一个batch数量的文本训练。其中，横坐标表示训练的全局步数，纵坐标表示关键词提取结果的$F1$值。可见，字向量过滤层可以将模型的$F1$值平均提升2%，有效改善了模型的过拟合现象。

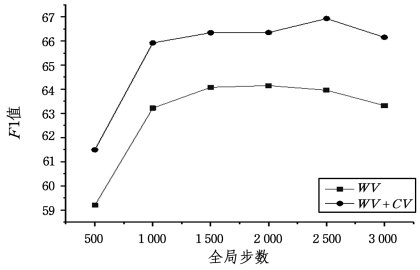

图6 WV模型与WV+CV模型的关键词提取效果的对比

Fig. 6 Comparision of keyword extraction effects between WV model and WV+CV model

实验表明，词向量的维度和隐藏层神经元节点数对实验效果的影响较明显。当选取80维以上的词向量时，模型的训练效果更为稳定；当隐藏层神经元节点数选取100以上时，模型的训练效果较好。与只使用词向量的模型相比，字向量与词向量相结合的模型的$F1$值可以提高2%～3%，模型的过拟合现象越显著，纠正效果就越好。因此，结合字向量模型预测结果，可以减轻模型的过拟合现象，改善关键词提取效果。

结束语 为了改进传统关键词提取算法在论文摘要关键词自动抽取任务上的效果，本文提出了一种基于注意力机制的关键词提取方法。在论文摘要数据集上的实验结果表明，该方法在准确率、召回率和F1值上优于其他方法。

本文结合词向量与字向量的序列标注模型，修正了模型的过拟合现象，并使用BiLSTM神经网络解决了远距离依赖问题；其引入注意力机制，与CRF层相比，可以更准确地提取文本内容特征，具有最高的$F1$值。该方法可用于多种短文

本数据集，具有较好的可扩展性。该模型只需要标注出文本中所包含的关键词即可进行训练，标注难度低，特征工程简单，不用考虑词性以及其他文本特征，且易于训练，收敛速度较快。但是，该模型只能提取数据集中已经存在的关键词，不能做到概括文本内容，而自动生成文本中可能不存在的合适的关键词，因此在语义理解方面有待进一步的研究。所提模型只考虑了从摘要中提取关键词，未考虑论文标题与关键词间的关系，如何结合论文标题共同提取关键词还有待进一步的研究。

参 考 文 献

[1] GOLLAPALLI S, CARAGRA C. Extracting Keyphrases from Research Papers Using Citation Networks [C]// Proceedings of the National Conference on Artificial Intelligence. Quebec: AAAI Press, 2014: 1629-1635.

[2] FLORESCU C, CARAGEA C. Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 1105-1115.

[3] HASAN K, NG V. Automatic keyphrase extraction: A survey of the state of the art[C]// Proceedings of the 27th International Conference on Computational Linguistics. Baltimore, Maryland, 2014: 1262-1273.

[4] LI G, WANG H. Improved automatic keyword extraction based on textrank using domain knowledge[C]// Proceedings of the 2014 Natural Language Processing and Chinese Computing. Berlin: Springer-Verlag, 2014: 403-413.

[5] BOUGOUIN A, BOUDIN F, DAILLE B. TopicRank: Graph-Based Topic Ranking for Keyphrase Extraction[C] // Proceedings of theInternational Joint Conference on Natural Language Processing. Nagoya, Japan, 2013: 543-551.

[6] TENEVA N, CHENG W. Salience rank: efficient keyphrase extraction with topic modeling[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 530-535.

[7] FLORESCU C, CARAGEA C. A Position-Biased PageRank Algorithm for Keyphrase Extraction[C]// Proceedings of the American Association for Artificial Intelligence. San Francisco: AAAI Press, 2017: 4923-4924.

[8] ZHANG C, WANG H, LIU Y, et al. Automatic keyword extraction from documents using conditional random fields[J]. Journal of Computational Information Systems, 2008, 4(3): 1169-1180.

[9] HADDOUD M, MOKHRARI A, LECROQ T, et al. Accurate Keyphrase Extraction from Scientific Papers by Mining Linguistic Information[C]// Proceedings of The Workshop on Mining Scientific Papers: Computational Linguistics and Bibliometrics. Istanbul, Turkey: CEUR-WS, 2015: 12-17.

[10] ONAN A, KORUKOGLU S, BULUT H. Ensemble of keyword extraction methods and classifiers in text classification[J]. Expert Systems with Applications, 2016, 57(3): 232-247.

[11] GOLLAPALLI S, LI X, YANG P . Incorporating expert knowledge into keyphrase extraction[C]// Processings of the American Association for Artificial Intelligence. San Francisco: AAAI Press, 2017: 3180-3187.

[12] ZHANG Q, WANG Y, GONG Y, et al. Keyphrase Extraction Using Deep Recurrent Neural Networks on Twitter[C]// Proceedings of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2016: 836-845.

[13] REKIA K, ZHANG Y, ZHANG W, et al. CCG Supertagging via Bidirectional LSTM-CRF Neural Architecture[J]. Neurocomputing, 2017, 283(12): 31-37.

[14] MOURAD G. Character-level neural network for biomedical named entity recognition[J]. Journal of Biomedical Informatics, 2017, 70(5): 85-91.

[15] ANDREJ Z, YORAM B, PASHA M, et al. Neural Named Entity Recognition Using a Self-Attention Mechanism[C] // Proceedings of International Conference on TOOLS with Artificial Intelligence. Boston: IEEE Computer Society, 2017: 652-656.

[16] SI Y, XIAO Y, XU J, et al. Recurrent neural network language model with vector-space word representations[C]// Proceedings of the International Conference on Learning Representations. Beijing: International Institute of Acoustics and Vibrations, 2014: 3024-3031.

[17] SUNDERMEYER M, SCHLUTER R, NEY H. LSTM Neural Networks for Language Modeling[C]// Proceedings of the 13th Annual Conference of the International Speech Communication Association Interspeech. Portland, OR, 2012: 194-197.

[18] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Nrural Networks, 2005, 18(5): 602-610.

[19] FENG S, LIU S, YANG N, et al. Improving attention modeling with implicit distortion and fertility for machine translation [C]// Proceedings of 26th International Conference on Computational Linguistics. Osaka, Japan, 2016: 3082-3092.

[20] TAN Z, WANG M, XIE J, et al. Deep Semantic Role Labeling with Self-Attention[C]// Proceedings of the American Association for Artificial Intelligence. San Francisco: AAAI Press, 2017: 4923-4924.

YANG Dan-hao, born in 1994, master. His main research interests include natural language processing.

FAN Chun-xiao, born in 1962, professor. Her main research interests artificial intelligence and internet of things.