基于神经网络的连动句识别

孙超 [1] ，曲维光 [1,2] ，魏庭新 [2,3] ，顾彦慧 [1] ，李斌 [2] ，周俊生 [1]
(1. 南京师范大学 计算机科学与技术学院，江苏省 南京市 210023
2. 南京师范大学 文学院，江苏省 南京市 210097
3. 南京师范大学 国际文化教育学院，江苏省 南京市 210097)

摘要

连动句是具有连动结构的句子，是汉语中的特殊句法结构，在现代汉语中十分常见且使用频繁。连动句语法结构和语义关系都很复杂，在识别中存在许多问题，对此本文针对连动句的识别问题进行了研究，提出了一种基于神经网络的连动句识别方法。本方法分两步：第一步，运用简单的规则对语料进行预处理；第二步，用文本分类的思想，使用 BERT 编码，利用多层 CNN 与 BiLSTM 模型联合提取特征进行分类，进而完成连动句识别任务。在人工标注的语料上进行实验，实验结果达到 92.71% 的准确率，F1 值为 87.41%。

关键词： 连动句 ； 文本分类 ； 神经网络

Recognition of serial-verb sentences based on Neural Network

SUN Chao[1], QU Weiguang[1,2], WEI Tingxin[2,3], GU Yanhui[1], LI Bin[2], ZHOU Junsheng[1]
(1.School of Computer Science and Technology,Nanjing Normal University,Nanjing, Jiangsu 210023,China;2.School of Chinese Language and Literature,Nanjing Normal University,Nanjing,Jiangsu 210097,China;3.International College for Chinese Studies, Nanjing Normal University，Nanjing,Jiangsu 210097,China)

Abstract

Serial-verb sentence is a sentence with several coordinated verbs in it. As a special syntactic structure it is very common and frequently used in modern Chinese. The grammatical structure and semantic relationship of serial-verb sentences are very complicated, which brings obstacles in its automatic recognition. This paper focuses on the recognition of Serial-verb sentence and proposes a recognition model based on neural networks. This method is implemented in two steps: the first step is to use rules to preprocess the corpus; the second step is to use BERT, the multi-layer CNN and the BiLSTM model to jointly extract features for classification, and then complete the sentence recognition task. Experimental results show that our model performs good in serial-verb sentence recognition,and the accuracy reaches 92.71% accuracy, while the F1 value reaches 87.41%.

Keywords: Serial-verb sentence , text classification , neural network

收稿日期： ；**定稿日期：**
基金项目：国家自然科学基金“汉语抽象意义表示关键技术研究”(61772278)；江苏省高校哲学社会科学基金“面向机器学习的汉语复句语料建设研究”(2019JSA0220); 国家社会科学基金“中文抽象语义库的构建及自动分析研究”（18BYY127）。

1 引言

通常情况下，动词是句子理解的关键，对句子的理解一般从动词入手。现代汉语中动词连用的现象大量存在，但相似形式所代表的语法结构和语义结构却不一定相同，有时甚至千差万别。在句子级别的语义研究方面，谓词所处的事件框架中包含的各种语义关系就构成了句子的语义结构。连动句包含多个谓词，蕴含了十分丰富的知识。连动句表示的是多个事件, 这些事件相互依赖并呈现出语义上的方式、顺承、目的、因果等关系，相互依赖影响并产生相互关联的事件 (刘雯旻, 2017)。因此获取连动句的方法将在自然语言理解领域中发挥重要的作用，有效的连动句识别可以在大规模语料中获取其中的连动句，从而便于对连用的动词或动词性短语之间的语义关系进行研究，有助于自然语言处理中句子级别的语义分析任务的研究和句法解析任务的研究，获取连动句的方法将在常识获取、智能网页等人工智能应用领域中发挥重要的作用。同时也为其他特殊句式的获取和处理提供了思路，从而帮助人们更加深入地理解自然语言。

连动句是汉语中一种较特殊的句式结构，在汉语中非常普遍。对连动句的研究可以从马建忠的《马氏文通》(成书于 1898 年) 中找到最早的踪迹。随后，又有很多语言学家都对连动句作了深入研究，其中，赵元任（《北京口语语法》）、张志公（《汉语语法常识》）、丁声书等（《现代汉语语法讲话》）、吕叔湘（《现代汉语八百词》）等都曾对连动结构做过分析和界定 (洪淼, 2004)。一般认为，连动句是指句中谓语为连动谓语的句子，即这个句子的谓语是两个或两个以上的动词连用，这些动词之间没有联合、动宾、偏正或主谓等关系，也没有明显的语音停顿，不用关联词语，而且这些动词都由同一个主语发出 (韩志玲, 倪蓉, 2012)。

综合前人的研究 (许有胜, 2007; 彭国珍 et al., 2013; 洪淼, 2013; 陈波 et al., 2013)，本文将所研究的连动句定义如下：在一个单句中，含有两个或两个以上的动词（或动词结构）且动词的施事为同一对象。其中第一个动词（简称 V1）的主语位置出现的名词短语 NP1 位置固定，而第二个动词结构（简称 V2）的名词短语 NP2，与 V1 的 NP1 同形，且必须隐含。其句法格式为：NP1+V1+(NP2[NP1])+V2，且 V1 和 V2 之间在语义上具有时序、目的、方式和原因等关系。本文所研究的连动句是形如“我去图书馆看书”，此句中的“去”和“看书”两个动词的施事都是“我”且只出现第一个动词的主语位置，“去”和“看”具有时序关系，且二者皆为动作行为动词，而像“地面人员看到丁毅找准跑道”一句中也含有两个动词“看到”和“找准”，但两者的施事分别是“地面人员”和“丁毅”，所以此句属于本文定义的非连动句。

抽象语义表示（abstract meaning representation, AMR）是近年来新兴的一种句子级的语义表示方法，突破了传统的句法树结构的限制，将一个句子语义抽象为一个单根有向无环图，很好地解决了论元共享的问题 (曲维光 et al., 2017)。而在连动句中存在着内部概念节点论元共享的现象，即在单句中多个谓词共享同一论元角色，这种 V1、V2 间的施事主语共享现象正是连动句区别于其他特殊句法结构的最主要特征，AMR 会将缺省的论元进行补全，得到完整语义表示。基于此特征可以从 AMR 图中获取可能的连动句，再经过人工校对得到连动句集合和非连动句集合，故本文选取了李斌等 (2017) 设计建立的中文 AMR 语料作为部分实验数据集。

本文的数据集主要有两部分组成，除使用抽象语义表示（AMR）体系标注的人教版小学 1-6 年级语文教材外；另一部分是清华树库的语料，经过人工标注赋予每个句子是连动句或非连动句的标签。

本文提出的连动句的识别方法分为两步：第一步：首先利用简单的规则剔除掉语料中一部分非连动句；第二步：基于神经网络做文本分类，将连动句与非连动句看作两个类别的文本进行分类，使用 BERT 编码，利用多层 CNN 和 BiLSTM 模型联合提取特征，进行句子分类，标签为“连动句”的文本，即为模型识别出的连动句。在此方法中，不需要手工筛选复杂特征，同时降低了对 NLP 领域的前置知识的需求。文本分类的实验效果达到 92.71% 的准确率，连动句识别的 F1 值为 87.41%。同时本工作也可以帮助 AMR 标注工作，定位连动句的位置，在后续工作中完成连动句中连动词和主语的识别以及连动词的语义关系识别，即可实现连动句的 AMR 自动标注。

2 相关工作

近年来针对连动句的研究主要集中在连动句对外教学的研究以及从汉语言文学角度研究分析连动句的句法和语义问题，针对连动句识别的研究工作较少。

许有胜 (2013) 提出了连动结构的自动识别和分析的方法。他主要从形式特征和语义角色两个方面编制出一些规则，尝试对连动结构进行自动识别和分析。但是由于连动结构的复杂性，

所设置的规则并不能涵盖所有情况，使得他提出的自动识别方法在很多环节的处理都存在问题，但他提供了一种基于“规则识别”的思路。

刘雯旻, 张晓如 (2017) 提出了一种基于规则和统计的连动句识别方法，他们构建了基于连动句形式特征和语义角色的基础规则库和被动名词库，利用互信息计算谓语动词与主语候选项的搭配强调，从而达到连动句识别的目的，实验结果达到 79.42% 的准确率,F1 值为 70.83%。

随着深度学习的发展，神经网络在解决文本信息处理相关任务中取得了较大的进展。本文将连动句的识别问题视为文本分类问题，提出一种基于神经网络的识别方法。深度学习的文本分类方法需要将文本输入到一个深度网络中，得到文本的表示形式，然后将文本表示形式输入到 softmax 函数中，得到每个类别的概率。目前，利用神经网络进行文本分类已经取得很多进展。Kim 最早提出将 CNN 应用于文本分类任务 (Kim Y, 2014)，Lai 等提出了 RCNN 模型，更好地利用了上下文信息 (Lai S et al., 2015)，Conneau 等在此基础上提出 VDCNN 模型，采用了深度卷积网络方法 (Conneau A et al., 2017)。Liu 等针对文本多分类任务提出基于 RNN 的不同共享机制模型 (Liu P et al., 2016)，Wang 等提出了 DRNN 模型，通过固定信息流动的步长提高文本情感分析的准确率 (Wang B, 2017)。

虽然许多神经网络的模型在文本分类任务中都取得了不错的表现，但连动句与非连动句的分类又与一般的文本分类任务不同，许多形式上十分相似的句子很可能不属于同一类别，更需要关注句中的动词间的语义关系和它的施事。基于此，本文利用 BERT 得到文本的表示形式，在训练过程中可以根据上下文动态的调整词向量，将其与多层 CNN 和 BiLSTM 进行组合作为连动句识别的模型，在人工构建的语料库上取得不错的效果。

3 模型设计

本模型采用 BERT 的输出结果作为字表示，将 BiLSTM 与两层 CNN 获取的局部特征相结合，用 Concatenate 连接，再经过一个全连接层，最后通过 softmax 层输出最终的判断结果，整体模型图如图 1所示。

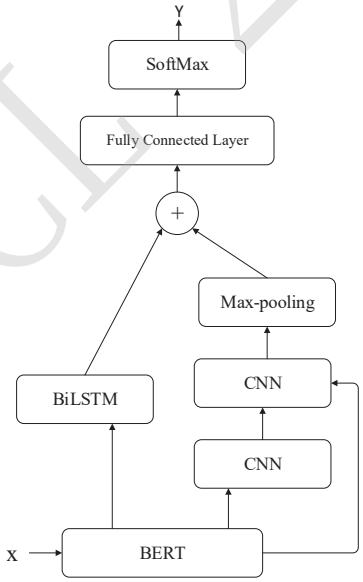

Figure 1: 模型设计

3.1 文本表示

BERT 是基于 Transformer 的双向编码器表示（Bidirectional Encoder Representation from Transformers） (Devlin J et al., 2018), 旨在通过联合调节所有层中的上下文来预先训练深度双向表示。使用双向的 Transformer 进行编码，使得在处理每个词的表示时都要考虑上下文信息，具体模型图如图 2所示。同时 BERT 预训练过程中使用了 Masked LM 和 Next Sentence Prediction 两种方法，迫使模型更多地依赖于上下文信息去预测词汇和句子，并且赋予了模型一定的纠错能力，分别捕捉词语和句子级别的 representation。

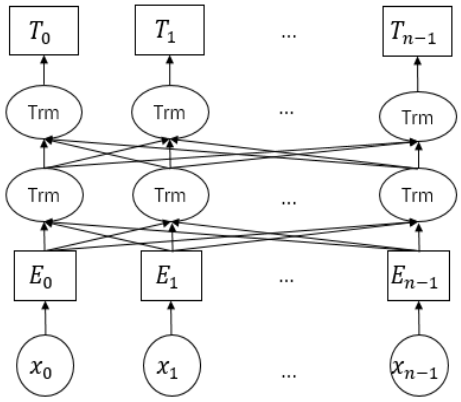

Figure 2: BERT 模型图

BERT 模型的输入不仅仅是字本身，它由三个 embeddings 向量相加而成，包含更多信息，输出则是已经融入全句语义的各个字的向量表示。BERT 输入向量表示示意图如图 3所示，Token Embedding 层将各个字转换成固定维度的向量，在 BERT 中每个字都会被转换成 768 维向量表示。BERT 能够处理输入句子对，segment embedding 层的作用是区分两个句子，前一个向量将 0 赋给第一个句子中的各个 token，后一个向量是把 1 赋给第二个句子中的各个 token。在本文中，输入的是一个句子，所以 segment embedding 全为 0。Position Embedding 实现编码序列的顺序性，当一个句子同一个字出现多次时，position embedding 提供了不同的向量。最终对“我去图书馆看书”一句将得到 3 个维度为（1,9,768）的向量，3 个向量按位相加最终得到大小为（1,9,768）的合成表示，富含更加丰富的语义信息。

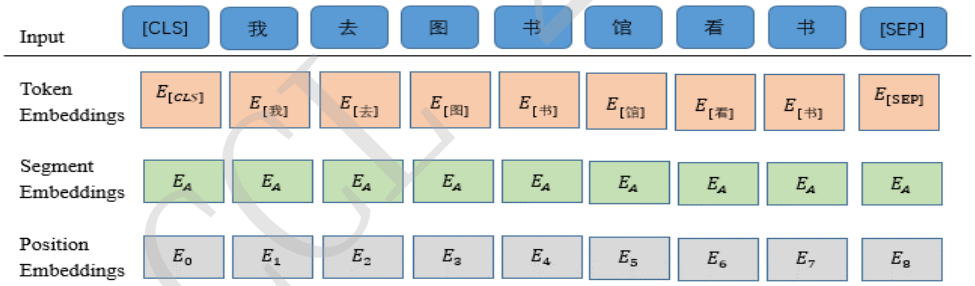

Figure 3: BERT 输入向量表示

3.2 特征提取

3.2.1 BiLSTM 层

将待判断的单句进行字级别的编码后将结果送入 BiLSTM 层，BiLSTM 将利用字在句子中的前后顺序，同时还可以捕获双向的较长距离的语义依赖关系，从而更好地判断当动词间相隔较远情况下是否可以构成连动关系。

BiLSTM 可以学习输入词的前后信息，从而有助于分类。给定由 n 个字组成的句子 X，它表示为一组向量 $(x_0, x_1, ..., x_{n-1})$，通过公式组（1）-（5）计算每个时间 t 的 LSTM 单元 (Chung J et al., 2018)。

$$i_t = \sigma(W_{x_i} \cdot x_t + W_{h_i} \cdot h_{t-1} + W_{c_i} \cdot c_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(W_{x_f} \cdot x_t + W_{h_f} \cdot h_{t-1} + W_{c_f} \cdot c_{t-1} + b_f) \tag{2}$$

$$c_t = (1 - i_t) \bigodot c_{t-1} + i_t \bigodot tanh(W_{x_c} \cdot x_t + W_{h_c} \cdot h_{t-1} + b_c) \tag{3}$$

$$o_t = \sigma(W_{x_o} \cdot x_t + W_{h_o} \cdot h_{t-1} + W_{c_o} \cdot c_t + b_o) \tag{4}$$

$$h_t = o_t \bigodot tanh(c_t) \tag{5}$$

其中 x_t，h_{t-1}，c_{t-1} 表示输入，h_t 和 c_t 表示输出。i_t、o_t、f_t 分别表示输入门、输出门和遗忘门。c_t 表示记忆单元向量。W_i、W_o、W_c 分别表示输入词向量 x_t，隐藏层状态 h_t 和记忆单元 c_t 的权重矩阵，b_i、b_f、b_c 和 b_o 分别表示偏差向量。$\bigodot$ 表示按位乘操作，σ 表示 sigmoid 激活函数。

通过 LSTM 可以得到与句子长度相同的隐层状态序列 $\{h_0, h_1, ..., h_{n-1}\}$，将前向 LSTM 与后向的 LSTM 结合成为 BiLSTM。通过对正向的时间序列和反向的时间序列进行训练，使输出的数据包含上下文的信息。解决了 LSTM 网络缺乏对上下文的联系，从而使模型获取更多的上下文信息 (Mike Schuster and Kuldip K Paliwal, 1997)。本文中使用的是两个 BiLSTM 堆叠形成的模型，中间使用一个全连接层进行降维，前一层 BiLSTM 的输出作为下一层 BiLSTM 的输入。

3.2.2 CNN 层

卷积神经网络是神经网络中提取局部特征的一种网络 (Wang J et al., 2017)，具有强大的特征学习和表示能力，基本结构由四层构成，分别是输入层、卷积层、池化层、全连接层。本模型中使用了两层 CNN 网络进行串联，第一层 CNN 的输入使用 BERT 的输出，将第一层 CNN 的输出和 BERT 的输出进行拼接作为第二层的输入。卷积层本质上是特征提取器，输入经过过滤器进行卷积操作后得到新的特征。设滤波器 $W \in \mathbb{R}^{m \times n}$，卷积得到：

$$y_{i,j} = \sum_{u=1}^{m} \sum_{v=1}^{n} w_{uv} x_{i-u+1, j-v+1} \tag{6}$$

其中，$m \ll M$，$n \ll N$。另外在卷积的标准定义基础上，根据不同任务的需求，可调整滤波器的滑动步长 (stride)、引入零填充 (zero padding) 来增加卷积的多样性，更灵活地进行特征抽取。两层滤波器采用不同的滑动步长，获取不同尺度的局部特征信息，有利于捕获句中不同距离动词间的信息。卷积层 (Wang S et al., 2018) 通过局部连接大大减少了网络参数的数量，通过权重共享使特征提取与数据位置无关，但其输出的神经元个数并没有显著地减少，容易造成过拟合，所以在卷积层之后再加上一个池化层，使用 Max-pooling 进行降维，同时增加平移不变性，模型更关注是否存在某些特征而非其位置，使得网络对一些细小的局部形态改变保持不变性，在减少数据量的同时保留有用的信息，最终得到固定长度的输出。

3.3 分类预测

将经过 CNN 层获得的特征与经过 BiLSTM 层获得的特征进行拼接，通过全连接层将特征整合到一起，同时对网络进行 Dropout 处理，以防止过拟合，随后送入 Softmax 进行预测。由于在数据集中连动句与非连动句分布不均，连动句数量较少，在计算损失函数时使用加权损失函数，使模型更多的关注样本数较少的类，更有利于模型识别连动句。

4 实验设置

4.1 连动句识别流程

本文对连动句识别流程如图 4所示，第一步，首先需要对语料进行切分操作，本文对连动句的识别是以单句为单位，以标点符号“。”、“，”、“：”、“；”为切分依据；之后再对语料进行词性标注工作。本文中使用的词性标注器是使用清华树库语料作为原始语料、用 BiGRU-CRF 模型自行训练所得，该模型可以实现对动词的细分类。本文采用范晓的《汉语动词概论》的分类系统 (范晓, 1980)，动词可根据它的表义功能分为动作行为动词、心理动词、使令动词，存现动词、判断动词、能愿动词、趋向动作动词、先导动词。动作行为动词和心理动词可充当连动

句中的 V。像“科长的口袋一下子鼓了起来”一句中，“起来”被标注为趋向动作动词，所以本文认为该句中只含有一个动词“鼓”。更加精准的词性标注，有利于语料在进行规则筛选时预先识别出更多的非连动句。由于在真实语料中，连动句与非连动句所占比例相差很大，且大部

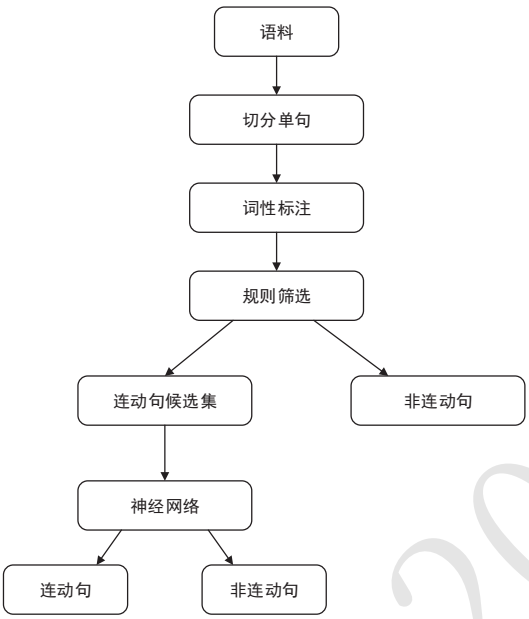

Figure 4: 连动句识别流程

分的简单非连动句（句中不含动词或只含有一个动词或动词结构）比较容易分辨，为了使神经网络模型将注意力放在学习与连动句在形式上比较相似的非连动句上，更好地学习两者的特征，本文首先制定了相应规则对语料进行预处理，预先识别出部分非连动句，其余句子可能为连动句，作为连动句的候选集送入神经网络进行分类。

本文设计规则将预先筛选的以下三种条件的句子排除在候选集之外：

条件一：不含动词或只含有一个动词（或动词结构）的句子。

条件二：含有关联词的紧缩句或复句。例如“常海一进病房就大咧咧地坐到程信的病床上”，句中含有“一…就…”形似连动，但该句属于紧缩句。

条件三：只含有多个动词和虚词的句子。当一个长句经过标点切分为独立小句或有些小句充当标题时会只有动词和虚词构成，例如“竞争与冲突”、“搜索前进”缺乏上下文信息，无法判断动词的主语，进而无法判断其是否为连动句。

通过“第一步”简单规则的辅助，可识别出大量非连动句，使连动候选集中连动句与非连动句的占比差距大大缩小，从相差 14 倍多缩小到不到 3 倍，表 1展示了筛选前后连动句与非连动句的数据的变化。

	连动句	非连动句	连：非连
处理前	7200	103188	1:14.3
处理后	7200	17852	1:2.5

Table 1: 规则筛选处理对比表

经过规则筛选后得到的连动句候选集中除连动句外，还含有大量的非连动句。这些非连动句中有一部分是和连动句相差明显的句子，例如“我在天色微明时走到了杨柳镇”根据语义可以得到“微明”和“走到”的施事明显是不一致；但还有部分与连动句相似度很高的其他特殊句式，例如，兼语句“他帮助妇女摆脱贫困”，虽然此处“帮助”和“摆脱”的施事并不一致，但此类句子在句式上与连动句相似，有一定区分难度。又如含有被动语义的句子，在判断动词的施事时也可能遇到困难，例如“他被通知住院”，“通知”和“住院”的施事并不相同，但在“他被诊断宣判生命只剩 2 个月”中“诊断”和“宣判”的施事又是同一个，该句为连动句，这些相似度极大的句子给本文的工作带来困难，也是神经网络模型需要重点学习的内容。

4.2 评价标准

本实验采用精确率（Precision）、召回率（Recall）、F1 值（F1-measure）、准确率（Accuarcy）作为评价标准。其中，TP：正确分类中连动句个数；FP：错误分类中连动句个数；TN：正确分类中非连动句个数；FN：错误分类中非连动句个数。

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$P = \frac{2 \cdot P \cdot R}{P + R} \tag{9}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

4.3 实验数据

连动句中 V1 和 V2 有共同的主语，在传统的语义表示方法上，由于树结构的限制，一般只标注主语和核心谓词的关系，而连动结构中其他谓词与主语的关系则被隐含。但这种隐含的语义关系正是连动结构区别于其他特殊句法结构最重要的特点。AMR 将补充出句中省略或隐含的成分，以还原出较为完整的句子语义，弥补传统句法表示的严重缺陷。

例如：“小白兔连忙挎起篮子往家跑。”一句的 AMR 图示注如图 5，“挎”和“跑”的主语都是“白兔”，原句中第二个动词“跑”的主语被省略了，在 AMR 图中会将此缺省补全，同时它也表示出了两个动词间的语义关系，此例中两个动词间的语义关系为“temporal（时序）”。

```
x1_小 x2_白兔 x3_连忙 x4_挎 x5_起 x6_篮子
x7_往 x8_家 x9_跑
(x11 / temporal
    :arg1()  (x4_x5 / 挎-01
          :arg0()  (x2 / 白兔
                :arg0-of()  (x1 / 小-01))
           :arg1()  (x6 / 篮子)
          :manner()  (x3 / 连忙))
     :arg2()  (x9 / 跑-01
          :arg0()  (x2 / 白兔)
         :arg3(x7/往)  (x8 / 家))
)
```

Figure 5: AMR 标注示例

本文根据 AMR 图抽取出小学语文 1-6 年级课本中的连动句，剩余句子为非连动句，同时又对清华树库的语料进行人工标注，共计 40667 个完整的句子。将句子切分为独立小句后，得到 11 万句分句，其中 7200 个独立小句为连动句。经过“第一步”处理后，共计 25052 个独立小句进行“第二步”神经网络模型的实验，按照 6:2:2 的比例划分训练集、开发集和测试集。

4.4 实验参数设置

使用 BERT 的基础版本，网络层数设置为 12，隐藏层数设置为 768，Self-Attention Head 设置为 12。在 BERT 中要预先设置 max_seq_length 参数，未达到此长度的句子要做 padding 处理，而超过此长度的数据将会被截断，造成信息丢失。同时若此参数设置过大会占用大量内存空间。本实验主要参数设置如表 2所示。

参数名	参数值
句子最大长度	50
神经元丢弃率	0.5
学习率	0.0001
提前终止	3

Table 2: 模型参数设置

4.5 实验结果及分析

为验证本文提出的方法的有效性，本实验主要与以下几种目前流行的文本分类模型进行对比，实验结果如表 3所示。

(1) 基于规则和统计：刘雯旻, 张晓如在 2017 年提出，他们构建了基于连动句形式特征和语义角色的基础规则库和被动名词库，利用互信息计算谓语动词与主语候选项的搭配强度，在他们人工标注的数据集进行实验。

(2)FastText：利用简单的三层模型（输入层、单层隐藏层、输出层），根据上下文预测文本的类别 (Joulin A et al., 2016)。

(3)TextCNN：利用 CNN 来提取句子中的关键信息，先将文本分词做 embedding 得到词向量, 再将词向量经过一层卷积, 一层 max-pooling, 最后将输出外接 softmax 实现文本类别的预测。

(4)TextRNN：RNN 模型由于具有短期记忆功能, 它通过前后时刻的输出链接保证了“记忆”的留存，引入门控机制解决长期依赖问题，捕获输入样本之间的长距离联系。

(5)BERT：用 Transformers 作为特征抽取器的深度双向预训练语言模型，在许多自然语言处理任务有很好的表现。

由表 3不同文本分类模型进行连动句识别的结果可知，本文提出的模型在连动句与非连动句分类的任务上具有很好的效果。除基于规则和统计的方法使用作者标注的语料外，其余神经网络的模型均使用本文中介绍的利用简单规则筛选后的语料。对比结果发现，FastText 模型基本没能学习到连动句的特征，在本任务上的效果较差；TextCNN 和 TextRNN 的效果相差不大，但都表现的还不够理想；而 BERT 模型 F1 较之前的模型有较大的进步，通过进一步分析 BERT 模型识别错误的句子发现，BERT 模型对长句的识别效果比较差。“连动句”这种语言现象可以出现在任何领域，它关注的是动词与动词的发出者之间的关系，而不是整个句子的语义关系，而且与词序有关。本文提出的模型使用 BERT 编码使模型获得了更多的语义信息，BiLSTM 层可以提取上下文不同距离的语义化信息, 同时 CNN 可以获取局部的特征，将多种特征进行组合，从而完成对连动句与非连动句的区分。

模型	连动句			ACC
	P	R	F1	
基于规则和统计	75.48%	66.72%	70.83%	79.42%
FastText	40.30%	98.24%	57.15%	41.41%
TextCNN	66.70%	68.68%	67.68%	74.09%
TextRNN	81.39%	57.58%	67.45%	77.89%
BERT	79.94%	79.18%	79.56%	89.03%
本文	86.78%	88.04%	87.41%	92.71%

Table 3: 不同文本分类模型结果对比表

在时间消耗方面，本文所提出的模型的收敛速度很快，图 6和图 7展示了模型在开发集上的迭代次数与 loss 和 acc 的变化曲线, 由图像可知模型在迭代几轮后便可得到在开发集上效果最好的模型参数，之后 loss 值会发生小范围的波动，为防止模型训练造成过拟合的问题，所以在实验中设置了提前终止参数，并使用 dropout 使模型得到更好的泛化效果。

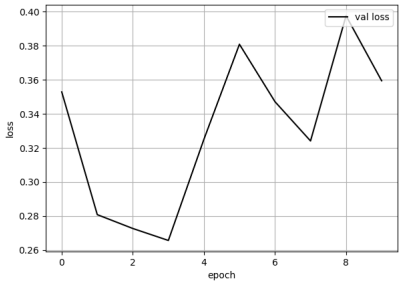

Figure 6: loss 曲线图

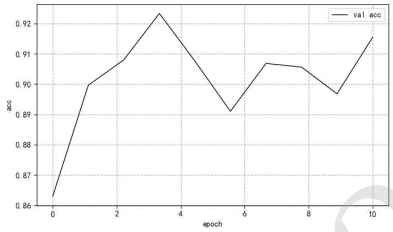

Figure 7: acc 曲线图

同时为了验证模型各层结构在实验中所起到的效果，特设置消融实验，结果如表 4 所示。由实验结果可知，使用随机初始化的 char 级别的词向量代替 BERT，实验的 F1 降低了约 20%，可见语义信息的获取和使用在连动句识别中起到了至关重要的作用，同时在训练过程中，本文提出的模型也会对词向量进行微调，以达到更好的表现。与此同时本文在获取特征时也采取了局部特征和全局特征组合的方式，更有助于连动句识别。通过实验发现，在使用 BERT+CNN 模型时，实验的 R 值较高，这说明模型可以尽量多将连动句挑选出来，将连动句误分为非连动句的情况较少，但同时它也将部分非连动句错误地识别为连动句，这是因为 CNN 侧重于提取句子局部信息，当句子局部出现两个动词或动词短语时，这一特征就会被 CNN 捕捉到，然而并非所有两个动词短语连用都是连动结构，可能是紧缩复句、兼语句、动词短语做宾语句等其他语言现象，因此造成这一模型 R 值较高而 P 值较低。而使用 BERT+BiLSTM 的模型则恰好不同，它的 P 值较高，BILSTM 侧重于捕捉句子全局信息，从整句角度去考察句子特征，因此很容易将复句、兼语句、动词宾语句等在句子层面排除在外，然而连动结构除了在句法层出现外，还可以以短语形式出现在各种句式结构的多个句法位置，BILSTM 模型对这类连动结构的识别能力较差，导致 R 值较低。本文使用的模型将两者的优势集中起来，提高了模型的 F1 值。

模型	连动句			ACC
	P	R	F1	
CNN+BiLSTM	60.80%	76.10%	67.60%	72.58%
BERT+CNN	75.71%	**92.90%**	83.43%	89.66%
BERT+BiLSTM	**88.33 %**	76.36%	81.91%	90.30%
BERT+BiLSTM+CNN	86.78%	88.04%	**87.41%**	**92.71%**

Table 4: 消融实验结果

根据实验结果我们发现连动句的识别错误分为两种，其一是非连动句错误识别为连动句，主要分为以下几种情况：

（1）汉语中一些动词的主语并非动词的施事者，导致模型判断出错，例如“出租车招手即停”一句中，句子的主语是“出租车”，但“招手”的施事是“人”而“停”的施事是“出租车”，二者并不相同。

（2）部分动词或动词短语做状语的状中结构和动词或动词短语做宾语的动宾结构识别易出错，它们在形式上与连动句相似，例如“宋东山心平气和地向小伙子笑笑”，此句中“心平气和”和“笑笑”的施事皆为“宋东山”，但该句为状中结构而非连动句；又如“川川总爱刨根问底”，“刨根问底”为动词，充当“爱”的宾语，且二者的施事都为“川川”，此句为动宾结构。

其二是某些连动句无法识别出来，主要分为以下两种情况：

（1）对多义词的识别效果不好，汉语中很多词语存在一词多义现象，但有些词语模型无法识别出它的动词义项，导致模型判断出错。例如“小刚看见这句话火了”, 火有多个义项，可以是名词、动词、形容词，但此处 embedding 未能准确表达出它为动词的语义。

（2）对长句的识别效果不好。例如“海淀区红山口甲 3 号国防大学医院疑难病研究中心的法集河使用近百味中药炮制膏药”模型识别为非连动句，但对“法集河使用近百味中药炮制膏药”可正确识别其为连动句。当句子某些修饰成分过长时，会影响模型的识别效果。

5 总结展望

本文根据连动句定义标注了连动句数据集，介绍了一种基于神经网络的连动句识别方法，先对语料进行切分和词性标注工作，再通过简单的规则进行第一轮非连动句的判断，之后使用 BERT 编码，将 BiLSTM 和 CNN 模型获取的特征进行组合，进行第二轮连动句与非连动句的判断，进而完成连动句的识别任务。实验表明，该模型取得了不错的识别效果。

本文的下一步工作是进一步提高连动句识别的准确率，同时对语料中识别出的连动句进一步找出其中的连动词，并识别它们之间的语义关系，从而帮助处理 CAMR 中连动句式的标注与解析工作。

参考文献

陈波, 姬东鸿, 吕晨. 2013. 基于特征结构的汉语连动句语义标注研究 [J]. 中文信息学报, 27(05):60-66+74.

范晓. 1980. 汉语的句子类型 [M]. 太原：书海出版社.

韩志玲, 倪蓉. 2012. 原型理论启发下的现代汉语连动句类型研究 [J]. 上海理工大学学报 (社会科学版), 34(01):41-45.

洪淼. 2004. 现代汉语连动结构研究 [D]. 南京师范大学.

洪淼. 2004. 现代汉语连动句式的语义结构研究 [J]. 西南民族大学学报 (人文社科版), 25(007):423-426.

李斌, 闻媛, 宋丽, 卜丽君, 曲维光, 薛念文. 2017. 融合概念对齐信息的中文 AMR 语料库的构建 [J]. 中文信息学报, 31(06):93-102.

刘雯旻. 2017. 基于汉语连动句的常识获取方法研究 [D]. 江苏科技大学,

刘雯旻, 张晓如. 2017. 一种基于规则和统计的连动句识别方法 [J]. 电子设计工程, 25(22):18-22.

彭国珍, 杨晓东, 赵逸亚. 2013. 国内外连动结构研究综述 [J]. 当代语言学, 15(03):324-335+378.

曲维光, 周俊生, 吴晓东, 戴茹冰, 顾敏, 顾彦慧. 2017. 自然语言句子抽象语义表示 AMR 研究综述 [J]. 数据采集与处理, 32(01):26-36.

许有胜. 2007. 连动结构研究综述 [J]. 兰州学刊, （09）：137-142.

许有胜. 2013. 连动结构的自动识别和分析 [J]. 巢湖学院学报, 15(04):108-115.

Chung J , Gulcehre C , Cho K H ,Bengio Y. 2018. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. In *arXiv preprint arXiv,1412.3555.*

Conneau A , Schwenk H , Barrault L , Yann Lecun. 2017. Very Deep Convolutional Networks for Text Classification[C]// In *Proceedings of the 15th Conference of the European Chapter of the Association for Compu-tational Linguistics EACL*,pages 1107-1116.

Devlin J, Chang M-W, Lee K, et al. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]// In *The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*,pages 4171-4186.

Joulin A , Grave E , Bojanowski P , et al. 2016. Bag of Tricks for Efficient Text Classification[C]// In *Proceedings of the fifty-fourth Annual Meeting of the Association for Computational Linguistics*,pages 427-431.

Kim Y. 2014. Convolutional Neural Networks for Sentence Classification[C]// In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*,pages 1746-1751.

Lai S, Xu L, Liu K,et al. 2015. Recurrent convolutional neural networks for text classification[C]// In *Proceedings of the twenty-ninth AAAI Conference on Artificial Intelligence*,pages 2267-2273.

Liu P, Qiu X, Huang X. 2016. Recurrent Neural Network for Text Classification with Multi-Task Learning[C]// In *Proceedings of the Twenty-fifth International Joint Conference on Artificial Intelligence*,pages 2873–2879.

Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural network [J]. In *IEEE Transactions on Signal Processing*,45(11):2673-2681.

Wang B. 2018. Disconnected Recurrent Neural Networks for Text Categorization[C]// In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*,pages 2311-2320.

Wang J, Wang Z, Zhang D, Yan J. 2017. Combining knowledge with deep convolutional neural networks for short text classification[C]// In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*,pages 2915-2921.

Wang S, Huang M, Deng Z. 2018. Densely Connected CNN with Multi-scale Feature Attention for Text Classification[C]// In *Proceedings of the Twenty-Seventh In-ternational Joint Conference on Artificial Intelligence*,pages 4468-4474.