

北京大学学报(自然科学版)  
Acta Scientiarum Naturalium Universitatis Pekinensis  
doi: 10.13209/j.0479-8023.2018.058

## 基于组合神经网络的语义省略“的”字结构识别

侍冰清<sup>1</sup> 戴茹冰<sup>2</sup> 曲维光<sup>1,2,3,†</sup> 顾彦慧<sup>1</sup> 周俊生<sup>1</sup> 李斌<sup>2</sup> 徐戈<sup>3</sup> 史胜旺<sup>1</sup>

1. 南京师范大学计算机科学与技术学院, 南京 210023; 2. 南京师范大学文学院, 南京 210097; 3. 闽江学院福建省信息处理与智能控制重点实验室, 福州 350121; † 通信作者, E-mail: wgqu\_nj@163.com

**摘要** 针对语义省略“的”字结构识别任务, 提出一种基于组合神经网络的识别方法。利用词语和词性, 通过双向 LSTM (long short-term memory) 神经网络, 学习“的”字结构深层次的语义语法表示。通过 Max-Pooling 层和基于 GRU (gated recurrent unit) 的多注意力层, 捕获“的”字结构的省略特征, 完成语义省略“的”字结构识别任务。实验结果表明, 所提模型在 CTB8.0 (Chinese Treebank8.0) 语料中, 能够有效地识别语义省略的“的”字结构, F1 值达到 96.67%。

**关键词** 神经网络; “的”字结构; 语义省略

### Hybrid Neural Network for Recognition of De Structure with Semantic Ellipsis

SHI Bingqing<sup>1</sup>, DAI Rubing<sup>2</sup>, QU Weiguang<sup>1,2,3,†</sup>, GU Yanhui<sup>1</sup>, ZHOU Junsheng<sup>1</sup>,  
LI Bin<sup>2</sup>, XU Ge<sup>3</sup>, SHI Shengwang<sup>1</sup>

1. School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023; 2. School of Chinese Language and Literature, Nanjing Normal University, Nanjing 210023; 3. Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou 350121; † Corresponding author, E-mail: wgqu\_nj@163.com

**Abstract** To solve the classification of the de structure containing the usage of semantic ellipsis, a hybrid neural network is built. Firstly, the network uses a bidirectional LSTM (long short-term memory) neural network to learn more syntactic and semantic information of the de structure. Then, the network employs Max-Pooling layer or GRU (gated recurrent unit) based multiple attention layers to capture features of ellipsis of the de structure by which the network can recognize the de structure containing the usage of semantic ellipsis. Experiments on CTB8.0 corpus show that the proposed approach can achieve efficient and accurate results.

**Key words** neural network; de structure; semantic ellipsis

作为现代汉语使用频率最高的虚词之一, “的”字用法灵活。在其用法中, 存在部分具有语义省略的“的”字结构, 如“开车的”, “我们应该做的”等, 这些不依附任何成分而独立充当名词性成分的“的”字结构通常在语义上伴有省略的成分。省略的“的”字结构是名词性偏正结构的语境变体<sup>[1]</sup>, 实质上是定中关系的偏正短语中心词隐去后的短语。其特征为词语后附着一个“的”字。然而, 并不是所有具有定中关系的偏正短语中心词都可以隐去, 从而形成“的”字结构。对于中心词可省的限制条件, 黄国营<sup>[2]</sup>和吕叔

湘<sup>[3]</sup>从语法角度分析了形如“X+的”结构中 X 与中心词的句法关系, 即当中心词为 X 的主宾语时, 中心词可省。孔令达<sup>[4]</sup>从意义的角度区分了 X 与中心词的语义类别关系, 并针对 X 是否具有区别性, 总结了一套形式化的鉴别方式。石毓智<sup>[5]</sup>从语言认知角度阐释了“的”字结构的形成机制。语义省略“的”字结构中隐含的成分对整体语义的理解有至关重要的作用, 正确识别具有语义省略的“的”字结构, 能够有效地减少因省略造成的语义自动理解障碍, 为补充句子完整的语义打下基础。

国家自然科学基金(61772278, 61472191)、江苏省高校哲学社会科学优秀创新团队项目(2017STD006)和福建省信息处理与智能控制重点实验室开放基金(MJUKF201705)资助

收稿日期: 2018-04-15; 修回日期: 2018-08-15; 网络出版时间: 2018-08-21 15:29:23

近几年,人们开始关注面向机器的“的”字结构研究。张坤丽等<sup>[6]</sup>构建了现代汉语广义虚词用法知识库(Chinese function word usage knowledge base, CFKB),其中包含助词“的”的虚词用法词典、规则库以及用法标注语料库。韩英杰等<sup>[7]</sup>在现代汉语词典、规则库、语料库“三位一体”的助词知识库基础上,采用基于规则的方法,研究“的”字用法的自动识别。刘秋慧等<sup>[8]</sup>在“三位一体”虚词用法知识库的基础上,分别采用基于规则、基于条件随机场和神经网络模型门循环单元,对助词“的”的用法进行自动识别。然而,鲜有研究者从语义成分省略的角度关注“的”字结构的语义完整性问题。仅从句法角度分类描述“的”字的用法和特征,不能深入地挖掘受语境和语言经济原则制约而省略的中心语语义,还原“的”字结构完整语义。鉴于上述背景,本文提出基于神经网络的语义省略“的”字结构自动识别方法。

近期,一种新型的句子语义表示方式——抽象语义表示(abstract meaning representation, AMR)受到研究者的广泛关注。AMR的主要设计目标在于描述和揭示句子中蕴含的完整的、深度的语义信息,以利于解决各种自然语言处理问题<sup>[9]</sup>。该方法突破了基于句法形式表示语义的传统方式,允许补充省略或隐含的语义概念,以便还原句子的完整语义。这种概念添加方式对汉语中的省略结构同样有良好的表示能力,能够较完整地补充省略成分。因此,AMR语料库构造和自动解析成为研究热点。李斌等<sup>[10-11]</sup>设计了中文AMR标注体系,并先后建立中文《小王子》AMR语料库和基于部分CTB8.0(Chinese Treebank-8.0)网络媒体真实语料的中文AMR(简称CAMR)语料库。我们对中文《小王子》AMR语料中需要添加的缺省概念进行统计分析,发现省略“的”字结构在所有省略结构中占有相当大的比例(45.7%)。

本文数据集由两部分组成,首先抽取CAMR语料库中所有的“的”字结构,并对添加概念的“的”字结构进行省略类别的自动标注,然后对CTB语料中未进行CAMR标注的部分,采用人工标注的方法获取更多省略“的”字结构。

本文在CTB8.0语料上探索基于神经网络的语义省略“的”字结构自动识别方法,将该问题视为二分类问题,提出一种基于组合神经网络的识别方法。首先,利用词语和词性,通过双向LSTM(long short-term memory)神经网络学习“的”字结构深层次的语

义语法表示。然后,通过Max-Pooling层和基于GRU(gated recurrent unit)的多注意力层,捕获“的”字结构的省略特征,完成语义省略“的”字结构识别任务。实验结果表明,在CTB8.0语料中,本文提出的模型能够有效地识别出语义省略的“的”字结构。

## 1 相关工作比较

韩英杰等<sup>[7]</sup>探讨了基于规则的助词“的”的用法自动标注,借助《现代汉语词典》、《虚词词典》、《现代汉语八百词》和《语法信息词典》等,人工归纳、总结和提取规则,也使用直接从语料中提取规则的方法,将规则转化为正则表达式,从而把“的”字用法标注问题转化成字符串模式匹配问题来处理。韩英杰等<sup>[7]</sup>基于规则的方法中需要借助助词知识库,在实验中要额外考虑规则的细化、扩充和调序,费时费力;并且,“的”的用法有39个,使用灵活,语法特征不明显,对于某些用法,不易写出规则的形式化描述。刘秋慧等<sup>[8]</sup>在对基于规则的方法及基于条件随机场的方法进行初步探究后,将深度学习方法引入“的”字用法的自动识别中,通过设置前向输入窗口,利用GRU单元获取长距离特征,提高了“的”的用法自动识别性能。刘秋慧等<sup>[8]</sup>设置的输入窗口中包含“的”右侧两个词及词性的特征,这种考虑符合部分“的”字用法的规则,如“的”字下文紧接某些词时,其类型是唯一的。但是,对某些不符合这个情形的“的”字用法,这种做法是一种干扰。刘秋慧等<sup>[8]</sup>认为GRU模型最后一个输出包含序列的全部有效特征,并将这个长距离特征直接作为分类的依据。

本文在以下4个方面有别于文献[7-8]的工作。1)从分类的角度看,文献[7-8]关注“的”字全部用法的分类,从句法角度描述“的”的用法和特征。但是,某些“的”字用法同时存在语义省略和非省略现象,文献[7-8]没有研究“的”字结构省略识别问题,而本文从语义成分省略角度关注“的”字结构。2)在语料选择上,文献[7-8]基于《人民日报》1998年1月和2000年4月的语料,人工标注“的”字用法。本文首先基于CAMR语料库,自动抽取所有“的”字结构,并对添加概念的“的”字结构进行省略类别的自动标注,然后对CTB语料中未进行CAMR标注的部分,采用人工标注的方法,获取更多省略“的”字结构。3)在方法上,韩英杰等<sup>[7]</sup>采用基于规则的方法,刘秋慧等<sup>[8]</sup>分别采用基于规则、基于条件随机场

和神经网络模型门循环单元的方法, 本文则提出一种基于组合神经网络的识别方法。4) 在模型上, 刘秋慧等<sup>[8]</sup>认为 GRU 模型最后一个输出单元包含序列的全部有效特征, 并将这个长距离特征直接作为分类的依据, 本文则将识别任务分成两个步骤。首先, 在内存模块, 通过双向 LSTM 神经网络学习“的”字结构深层次的语义语法表示。然后, 在抽取模块, 通过基于 GRU 的多注意力层和 Max-Pooling 层, 捕获“的”字结构的省略特征, 完成语义省略“的”字结构识别任务。

## 2 基于组合神经网络的语义省略“的”字结构识别

本文采用组合神经网络模型, 包含 4 个部分: 输入模块、内存模块、抽取模块和输出模块, 网络结构如图 1 所示。对于输入的“的”字结构  $s = \{s_1, \dots, s_{de}, \dots, s_N\}$ , 其中  $s_i = (w_i, p_i), 1 \leq i \leq N$  包含  $i$  位置上词语  $w_i$  及其词性  $p_i$ , 本文旨在识别围绕  $s_{de}$  构成的“的”字结构是否省略。在前 3 个模块, 对“的”字结构的词语和词性分别进行相同操作, 在输出模块中合并两者的处理结果。下面从词语的视角介绍前 3 个模块。

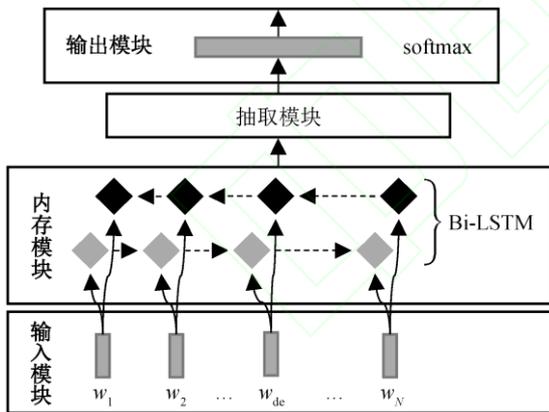


图 1 语义省略“的”字结构识别模型结构  
Fig. 1 Model architecture for recognition of the de structure with semantic ellipsis

### 2.1 输入模块

通过非监督学习方法, 比如 Skip-gram<sup>[12]</sup>, 得到一个词语向量查询表  $L \in \mathbb{R}^{d \times |V|}$ ,  $d$  是词语向量的维度,  $|V|$  是词语词汇表的长度。在输入模块中, 根据

输入序列  $w = \{w_1, w_2, \dots, w_N\}$ , 对  $L$  进行查询, 得到对应的词嵌入表示序列  $v = \{v_1, v_2, \dots, v_N\}$ , 其中  $v_i \in \mathbb{R}^d, 1 \leq i \leq N$ 。若词语向量查询表中不存在当前词语, 则采用特殊符号“UNK”的向量表示, “UNK”随机初始化。

### 2.2 内存模块

Tang 等<sup>[13]</sup>在 MemNet 中, 简单按词语在原句子中的顺序拼接词语的向量, 完成内存模块的构建。然而, 韩英杰等<sup>[7]</sup>在基于规则的方法中发现, 即使借助助词规则库, 在应用中也需要考虑规则之间的调序。这意味着, 达到人类标注水平的规则必定是非常复杂的, 应当使用一些循环神经网络方法, 使得内存模块含有短语级别的信息<sup>[14]</sup>, 比如在很多自然语言处理任务中表现优秀的单层或多层 Bi-LSTM<sup>[15]</sup>。Sachan 等<sup>[16]</sup>发现, 与卷积神经网络相比, Bi-LSTM 具有较好的获得语义信息的能力。因此, 本文选择 Bi-LSTM 来获得“的”字结构深层次的语义语法信息, 词性输入采用双层 Bi-LSTM, 词语输入则采用单层 Bi-LSTM。

词嵌入表示序列  $v = \{v_1, v_2, \dots, v_N\}$  作为 Bi-LSTM 的初始输入, 经过模型编码, 获得“的”字结构的抽象表示。其中, 对于词语向量  $v_t (1 \leq t \leq N)$ , 在第  $l$  层 Bi-LSTM 中, 通过该层前向 LSTM 计算, 获得向量表示  $\vec{h}_t^l$ , 通过后向 LSTM 计算, 获得向量表示  $\overleftarrow{h}_t^l$ 。如果共  $L$  层 Bi-LSTM, 最后将得到内存模块  $M^* = \{m_1^*, m_2^*, \dots, m_N^*\}$ , 内存片段  $m_t^* = [\vec{h}_t^L; \overleftarrow{h}_t^L], 1 \leq t \leq N$ 。[ ; ] 表示各向量连接构成一个向量, 也就是将第  $L$  层前向 LSTM 和后向 LSTM 的输出结果进行连接。

上述内存模块通过循环神经网络保留词序信息, 对“的”字结构进行深层语义语法建模, 但忽略了词语与“的”之间的相对位置信息。Chen 等<sup>[14]</sup>指出, 较好的距离公式应该使用两个词在依存树中的距离, 但是本文希望该任务仅仅基于数据驱动。因此, 对于  $i$  位置的词语与“的”字的距离, 采用 Chen 等<sup>[14]</sup>提出的公式:

$$u_i = \frac{i - \text{de}_{\text{index}}}{i_{\text{max}}}, \quad (1)$$

$$\text{weight}_i = 1 - |u_i|, \quad (2)$$

其中,  $u_i (1 \leq i \leq N)$  是词语与“的”字的距离,  $\text{de}_{\text{index}}$  是“的”字所在位置,  $i_{\text{max}}$  是“的”字结构的长度,  $\text{weight}_i$

是位置权重。除此以外,借鉴 Chen 等<sup>[14]</sup>不仅隐舍地考虑位置权重,还在内存模块中显性地放入位置信息,最终得到内存模块  $M=\{m_1, m_2, \dots, m_N\}$ , 其中  $m_i = [\text{weigh} \odot m_i^* u_i; \quad ] \quad 1 \leq i \leq N$ ,  $\odot$  是元素级别的向量乘法。本文也考虑了 Tang 等<sup>[13]</sup>提出的 4 种距离公式,但是效果没有超过 Chen 等<sup>[14]</sup>的公式。

### 2.3 抽取模块

抽取模块包含两个步骤:第一步,从内存模块中提炼“的”字结构的省略特征;第二步,合理地组织省略特征。具体地,采取基于 GRU 的多注意力机制和 Max-Pooling 两种抽取模式。

#### 2.3.1 基于 GRU 的多注意力层的抽取模块

本节介绍通过多注意力层从内存模块中抽取多个省略特征,并使用 GRU 单元将其合理组织。

一般认为,随着网络层数的加深,模型的刻画能力更强,可以获得更抽象的特征信息<sup>[17]</sup>,对省略类型“的”字结构的识别更有帮助。多注意力层可以在不同的层中注意到输入的不同位置<sup>[18]</sup>。使用非线性操作去组合不同的注意力层结果,可以避免抽取的内容仅仅是内存模块的线性组合<sup>[14]</sup>。该抽取模块的结构如图 2 所示。

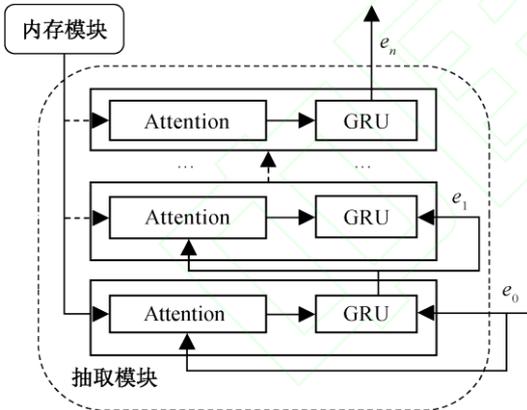


图 2 基于 GRU 的多层注意力层抽取模块

Fig. 2 Extraction module based on GRU and multiple attention layers

注意力层  $t$  的输入包括内存片段  $m_j (1 \leq j \leq N)$ 、GRU 的前一层隐藏状态  $e_{t-1}$ 、“的”字的词语向量  $v_{de}$  以及词语  $j$  与“的”字相对位置的距离  $u_j (1 \leq j \leq N)$ 。“的”字的词语向量和 GRU 的前一层隐藏状态可以指导计算当前注意力层中各内存片段的得分,有助于抽取与“的”字结构省略有关的信息。相对位置距离

有助于得到词语  $j$  与“的”的句法和语义关联。首先计算每一个内存片段  $m_j (1 \leq j \leq N)$  的注意力得分  $a_j^t$  以及该注意力层结果  $i_t$ :

$$g_j^t = W_{\text{att}} [m_j; e_{t-1}; v_{de}; u_j] + b_{\text{att}}, \quad (3)$$

$$a_j^t = \frac{\exp(g_j^t)}{\sum_{k=1}^N \exp(g_k^t)}, \quad (4)$$

$$i_t = \sum_{j=1}^N a_j^t m_j, \quad (5)$$

$W_{\text{att}}$  是权值参数,所有注意力层共享; $b_{\text{att}}$  是偏置项。根据 Tang 等<sup>[13]</sup>的经验,这里设置注意力层数为 3<sup>[13]</sup>。

接着,在注意力层之间应用 GRU 单元,其最后一个输出合理地组织了所有省略特征,将这个特征作为抽出模块的输出。GRU 主要通过“门”来更新状态,门可以让信息选择性地通过。GRU 单元中的非线性操作公式如下:

$$r = \text{sigmoid}(W_r i_t + U_r e_{t-1}), \quad (6)$$

$$z = \text{sigmoid}(W_z i_t + U_z e_{t-1}), \quad (7)$$

$$e' = \tanh(W_x i_t + W_g (r \odot e_{t-1})), \quad (8)$$

$$e_t = (I - z) \odot e_{t-1} + z \odot e', \quad (9)$$

$r$  是重置门,  $z$  是更新门,  $i_t$  是当前输入,  $e'_t$  和  $e_t$  表示候选激活状态和激活状态。与 LSTM 相比, GRU 具有更少的参数,更简单的结构,且更容易收敛<sup>[19]</sup>,所以这里选择 GRU。

#### 2.3.2 基于 Max-Pooling 的抽取模块

使用 Max-Pooling 提炼最重要的省略特征,抽取结构如图 3 所示。Max-Pooling 是元素级别的操作<sup>[20-21]</sup>,计算公式如下:

$$\text{extract\_de} = \max_{i=1}^N m_i. \quad (10)$$

由于语义省略“的”字结构识别是二分类任务,而 Max-Pooling 抽取的向量维度较大,因此本文再添加一层前馈神经网络(feedforward neural network, FNN),将其结果作为 Max-pooling 抽取模块的输出。

### 2.4 输出模块

在输出模块,本文连接基于词语和词性的“的”字结构省略特征,将其送入 soft max 层得到最终分类概率。计算公式如下:

$$p(d) = \text{soft max}(W_s [\text{extract\_de}_w; \text{extract\_de}_p] + b_s), \quad (11)$$

其中,  $p(d) \in \mathbb{R}^2$ ,  $W_s$  和  $b_s$  是分类层的参数。

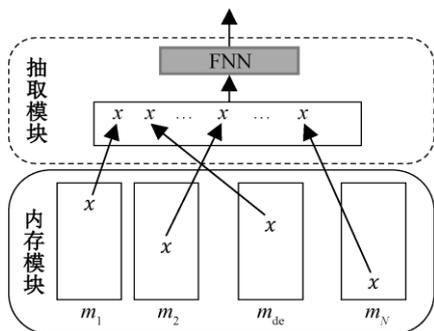


图3 基于 Max-Pooling 的抽取模块  
Fig. 3 Extraction module based on Max-Pooling

模型采用交叉熵损失函数, 计算公式如下:

$$\text{Loss} = \sum_{d \in D} \sum_{i \in C} y_i(d) \log p_i(d), \quad (12)$$

$C$  是“的”字结构的类型,  $D$  是训练数据集,  $p_i(d)$  是本文模型得到的样本  $d$  属于类型  $i$  的概率,  $y(d)$  是样本  $d$  的标签。我们用反向传播来计算参数的梯度, 神经网络模型优化选用 Adam 算法<sup>[22]</sup>, 正则化方法采用 dropout<sup>[23]</sup>。

### 3 实验

#### 3.1 语料

AMR 允许重新分析和补充概念, 能完整地表达一个句子的语义<sup>[10]</sup>。图4给出“跳舞的走了”的 AMR 表示。AMR 可以根据上下文添加概念 person, 作为“跳舞”的 arg0。同时, 中文 AMR 还会标出语义省略“的”字结构的虚词“的”。AMR 的这个特点解决了传统的句法表示方法无法应对的省略和词内分析困境<sup>[11]</sup>。

本文抽取 CAMR 语料库中所有的“的”字结构, 并对添加概念的“的”字结构进行省略类别的自动标

```
跳舞的走了
x3/走-02
:arg0 p/person
:arg0-of(x2/的) x1/跳舞-01
```

图4 概念补充的“的”字结构示例  
Fig. 4 Example of concept adding of the de structure

注。共有 11829 个“的”字结构, 其中非省略类型“的”字结构 11165 个, 省略类型“的”字结构 664 个。本文综合语法类词典《现代汉语八百词》<sup>[3]</sup>、现代汉语广义虚词用法知识库<sup>[6]</sup>及 CAMR 语料库, 分析省略与非省略类型“的”字结构中的上下文特征, 将“的”的意义分成 4 个义项, 详见表 1。

表 1 中, 义项 3 中全部“的”字结构和义项 4 中部分“的”字结构含有语义省略成分, 是本文讨论的省略类型“的”字结构。其余是非省略类型“的”字结构。可以看出, 两个类别的语料数量悬殊。因此, 对 CTB 语料中未进行 CAMR 标注的部分, 采用人工标注的方法获取更多省略类型“的”字结构, 以期平衡两者数量。最终, 省略类型“的”字结构增长到 1830 个。为了进一步平衡两种类型的数据量, 去除非省略类型“的”字结构对应的各义项近一半的数目, 数据情况详见表 2。

本文设置训练集和测试集的比例为 4:1。其中, “的”字的每种用法也按照该比例分发到训练集和测试集。关于模型初始的词语向量, 我们采用 word2vec 工具对 CTB 分词语料进行训练, 不随迭代更新; 词性向量随机初始化, 并在训练过程中更新。

#### 3.2 参数设置

实验中涉及多个超参数, 组合神经网络模型的超参数设置如表 3 所示。

表 1 “的”主要义项及用法描述

Table 1 Usages of auxiliary “DE” and their frequency in CAMR corpus

义项	频率	释义	用法	例句
义项 1	10189 (10189)	构成“的”字短语修饰名词	名 动 形 副 小句 介词短语+的+名	这些都和当今~社会现状何等相似。
义项 2	345 (345)	表示一定的语气	用在陈述句或疑问句末尾	他什么时候走~?
义项 3	473 (1277)	构成“的”字短语代替名词, 即省略“的”字结构”	名 动 形 小句+的	更令人回味~是贯穿始终的热闹。
义项 4	822 (1184)	构成“的”字短语做谓语(若前面有“是”, 构成“是...的”句式)	名 代 动 形 小句 四字语+的	回头想想, 人生的路, 好像都不是自己预期~。

说明: 括号内数字为获取更多省略类型“的”字结构后的各义项频率。

表 2 数据集  
Table 2 Datasets

语料	非省略“的”字结构个数	省略“的”字结构个数
Data1	11165	664
Data2	11165	1830
Data3	6191	1830

表 3 超参数设置  
Table 3 Settings of hyperparameters

序号	超参数名称	参数值
1	词语向量维数	100
	词性向量维数	20
	词语 Bi-LSTM 层隐藏单元向量维数	64
	词性 Bi-LSTM 层隐藏单元向量维数	32
	学习率	0.001
2	Max-Pooling 抽取模块输出向量维数	16
3	词语 GRU 隐藏单元向量维数	16
	词性 GRU 隐藏单元向量维数	20

目前,尚没有完善的理论可以自动且准确地界定“的”字结构的范围。本文参考刘秋慧等<sup>[8]</sup>基于条件随机场的自动识别方法,即提取向前、向后的词和词性等上下文环境,作为识别省略与非省略类型“的”字结构的依据。具体地,我们规定从“的”字前面 10 个词语起到“的”字后面 6 个词语止,构成一个“的”字结构,长度不足的补齐。

表 3 中,第 1 组超参数是输入模块和内存模块的参数,第 2 组是在 Max-Pooling 抽取模块中所需的参数,第 3 组是基于 GRU 的多注意力层的实验中所需的超参数。

### 3.3 实验结果与分析

#### 3.3.1 方法间的对比实验

韩英杰等<sup>[7]</sup>和刘秋慧等<sup>[8]</sup>关注“的”字全部用法的分类,只从句法角度描述“的”的用法和特征,其采用的某些“的”字用法仅属于省略类型“的”字结构,他们提出的方法可能在某种程度上也适用于本文的任务。因此,为了验证模型的有效性,基于本文数据集,我们参考刘秋慧等<sup>[8]</sup>基于条件随机场和神经网络模型门循环单元的方法,实现两个基线方法: CRF 和 GRU。CRF 使用刘秋慧等<sup>[8]</sup>提出的特征模板。对于 GRU,由于刘秋慧等<sup>[8]</sup>没有明确网络模型的超参数设置(如词语向量维度等),这里均采用与本文实验相同的超参数设置,并以“的”字位置对应的隐藏状态作为输出。这样做的结果比原模型中以序列的最后一个输出单元作为输出的实验效果好。

表 4 是与基线方法对比的实验结果,其中 ACC 是所有“的”字结构分类结果的正确率,非省略-F1 是非省略“的”字结构识别结果的 F1 值,省略-F1 是省略“的”字结构的识别结果的 F1 值。可以看出,CRF 的识别性能与语料规模成正比,但 GRU 在提升省略“的”字结构识别的同时,严重影响非省略类型的识别效果,而本文提出模型的性能均优于 CRF 和 GRU。

#### 3.3.2 模块分析

下面通过对比实验,分析各模块在模型中起到的作用。网络结构设置详情见表 5。其中, M1, M2 和 M3 分别为内存模块的 3 种设置; L1 和 L2 分别表示在内存模块中使用单层和双层双向 LSTM 神经网络; None 表示不使用抽取模块,直接连接“的”字位置上的前向、后向内存片段,将其视为“的”字结构省略特征,并送入输出层; Att 表示在抽取模块使用单

表 4 与基线方法对比实验结果  
Table 4 Performance of different methods

方法	Data1			Data2			Data3		
	ACC	非省略-F1	省略-F1	ACC	非省略-F1	省略-F1	ACC	非省略-F1	省略-F1
CRF	0.9387	0.9677	0.3830	0.9407	0.9644	0.7819	0.9451	0.9649	0.8739
GRU	0.9601	0.9793	0.4268	0.9143	0.9522	0.5873	0.8737	0.9209	0.6873
M3+GRU+Att	0.9877	<b>0.9935</b>	0.8929	0.9856	<b>0.9917</b>	0.9497	0.9838	0.9895	0.9642
M3+Max-pooling	<b>0.9885</b>	0.9934	<b>0.8996</b>	<b>0.9857</b>	<b>0.9917</b>	<b>0.9499</b>	<b>0.9850</b>	<b>0.9903</b>	<b>0.9667</b>

说明: 加粗数字为模型对比结果的最高值。下同。

层注意力机制,因为注意力层的结果本质上是内存模块的线性组合,如果多注意力层之间没有非线性操作,那么其最后结果还是内存模块的线性组合<sup>[14]</sup>,所以这里将多注意力层改为单注意力层;GRU+Att对应2.3.1节中基于GRU的多注意力层的抽取模块;Max-pooling对应2.3.2节中基于Max-Pooling的抽取模块。

图5显示,所有模型能够有效地分类“的”字结构,正确率都超过97.5%。图6显示,所有模型能够有效地识别非省略类型“的”字结构,F1值都超过98.5%。图6和7显示,随着语料扩充,省略与非省略类型“的”字结构的数据量逐渐趋于平衡,所有模型识别省略类型“的”字结构的性能越来越好,同时,识别非省略类型“的”字结构的性能没有受到明显的影响。

本文着重讨论省略类型“的”字结构的识别结果。图7显示,所有模型在数据集Data3上的表现都是最优的,因此接下来的实验均基于数据集Data3。

表5 网络结构设置

Table 5 Settings of networks

网络结构	词语内存模块	词性内存模块	抽取模块
M3+None	L1	L2	None
M3+Att	L1	L2	Att
M1+Max-pooling	L1	L1	Max-pooling
M2+Max-pooling	L2	L2	Max-pooling
M3+Max-pooling	L1	L2	Max-pooling
M3+GRU+Att	L1	L2	GRU+Att

表6是采用不同内存模块设置的对比实验结果,可以看出,M3+Max-pooling的性能超过M2+Max-pooling。虽然“的”字结构是一个复杂的结构,需要深层的模型在内存模块学习到其更本质的抽象表示,但根据助词知识库等研究成果中对“的”字用法的描述,“的”字结构的省略特征更依赖于句法信息。也就是说,从词语角度,词语包含语法和语义信息,可能

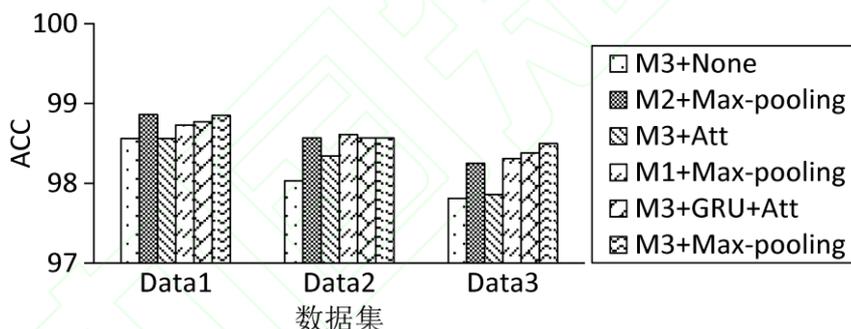


图5 “的”字结构分类结果

Fig. 5 Results of classification of the de structure

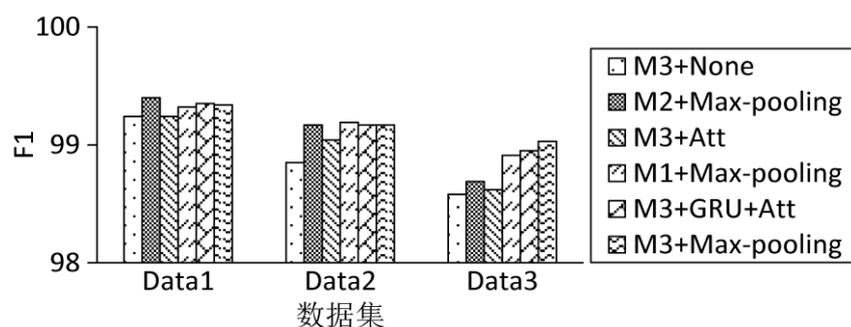


图6 非省略类型识别结果

Fig. 6 Results of recognition of the de structure without semantic ellipsis

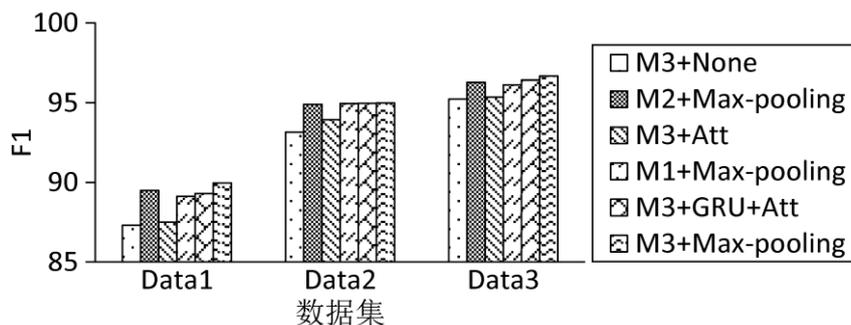


图 7 省略类型识别结果

Fig. 7 Results of recognition of the de structure with semantic ellipsis

不需要采用更倾向于学习语义信息的复杂模型, 比如两层 Bi-LSTM。从表 6 还可以看出, 模型 M3+Max-pooling 的性能超过 M1+Max-pooling。从词性角度来看, 词性本身不含语义, 仅提供语法信息, 因此可以适当地使用复杂模型学习其向量表示, 比如两层 Bi-LSTM。

表 6 内存模块对比实验结果

系统名称	P	R	F1
M1+Max-pooling	0.9589	0.9668	0.9628
M2+Max-pooling	<b>0.9613</b>	0.9613	0.9613
M3+Max-pooling	<b>0.9613</b>	<b>0.9720</b>	<b>0.9667</b>

表 7 是采用不同抽取模块设置的对比实验结果。注意力层的结果本质上是内存片段的线性组合, 也就是说, M3+None 的模型设置是 M3+Att 的特殊例子。但是, M3+None 的性能低于 M3+Att, 说明设立抽取模块是有必要的, 我们需要基于“的”字结构中所有的词来提取“的”字结构的省略特征。M3+GRU+Att 的 F1 值比 M3+Att 提高 1.06%, 说明合理地组织省略特征是有必要的。M3+GRU+Att 的性能暂时落后于 Max-pooling, 可能是由于基于注意力机制的网

表 7 抽取模块对比实验结果

系统名称	P	R	F1
M3+None	0.9407	0.9641	0.9523
M3+Att	<b>0.9641</b>	0.9432	0.9536
M3+Max-pooling	0.9613	<b>0.9720</b>	<b>0.9667</b>
M3+GRU+Att	0.9615	0.9669	0.9642

络拥有更多的参数, 其优化所需的语料规模更大。我们还可以在图 7 看到, 随着语料的扩充和不同类别数据量的平衡, M3+GRU+Att 与 M3+Max-pooling 的性能差距不断缩小。相信在语料更充足的情况下, M3+GRU+Att 会有更好的效果。目前, M3+Max-pooling 性能最佳, F1 值达到 96.67%。

#### 4 结语

本文从语义成分省略的角度关注“的”字结构的语义完整性问题, 研究了基于组合神经网络的语义省略“的”字结构识别方法。通过内存模块学习“的”字结构的深层次语义语法表示, 利用抽取模块提取和组织省略特征, 完成识别任务。据我们所知, 这是首次利用神经网络解决语义省略“的”字结构自动识别。实验结果显示, 该模型能够取得较好的识别效果。随着语料规模的扩展, 有可能取得更好的效果。

在下一步的研究工作中, 我们将针对省略类型的“的”字结构, 对其缺省的信息进行补全, 并尝试将补全的“的”字结构用于语义自动解析等工作中。

#### 参考文献

- [1] 徐阳春. 关于虚词“的”及其相关问题研究[D]. 上海: 复旦大学, 2003: 48-74
- [2] 黄国营. “的”字的句法、语义功能. 语言研究, 1982(1): 101-129
- [3] 吕叔湘. 现代汉语八百词. 北京: 商务印书馆, 1980: 156-163
- [4] 孔令达. “名\_1+的+名\_2”结构中心名词省略的语义规则. 安徽师大学报(哲学社会科学版), 1992, 20(1): 103-107
- [5] 石毓智. 论“的”的语法功能的同一性. 世界汉语教学, 2000(1): 16-27

- [6] 张坤丽, 咎红英, 柴玉梅, 等. 现代汉语虚词用法知识库建设综述. 中文信息学报, 2015, 29(3): 1-8
- [7] 韩英杰, 咎红英, 张坤丽, 等. 基于规则的现代汉语常用助词用法自动识别. 计算机应用, 2011, 31(12): 3271-3274
- [8] 刘秋慧, 张坤丽, 许鸿飞, 等. 助词“的”用法自动识别研究. 北京大学学报(自然科学版), 2018, 54(3): 466-474
- [9] 曲维光, 周俊生, 吴晓东, 等. 自然语言句子抽象语义表示 AMR 研究综述. 数据采集与处理, 2017, 32(1): 26-36
- [10] 李斌, 闻媛, 卜丽君, 等. 英汉《小王子》抽象语义图结构的对比分析. 中文信息学报, 2017, 31(1): 50-57
- [11] 李斌, 闻媛, 宋丽, 等. 融合概念对齐信息的中文 AMR 语料库的构建. 中文信息学报, 2017, 31(6): 93-102
- [12] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality // Proceedings of the 26th Annual Conference on Neural Information Processing Systems. Lake Tahoe, 2013: 3111-3119
- [13] Tang D, Qin B, Liu T. Aspect level sentiment classification with deep memory network // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, 2016: 214-224
- [14] Chen P, Sun Z, Bing L, et al. Recurrent attention network on memory for aspect sentiment analysis // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, 2017: 452-461
- [15] Schuster M, Paliwal K K. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681
- [16] Sachan D S, Zaheer M, Salakhutdinov R. Investigating the working of text classifiers [EB/OL]. (2018-01-19) [2018-06-29]. <https://arxiv.org/abs/1801.06261>
- [17] Sukhbaatar S, Szlam A, Weston J, et al. End-to-end memory networks // Proceedings of the 28th Annual Conference on Neural Information Processing Systems. Montreal, 2015: 2440-2448
- [18] Kumar A, Irsoy O, Ondruska P, et al. Ask me anything: dynamic memory networks for natural language processing // Proceedings of the 33rd International Conference on Machine Learning. New York City, 2016: 1378-1387
- [19] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [EB/OL]. (2014-12-11) [2018-06-29]. <https://arXiv.org/abs/1412.3555>
- [20] Conneau A, Kiela D, Schwenk H, et al. Supervised learning of universal sentence representations from natural language inference data // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, 2017: 670-680
- [21] Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning // Proceedings of the 25th International Conference on Machine Learning. Helsinki, 2008: 160-167
- [22] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2014-12-22) [2018-06-29]. <https://arxiv.org/abs/1412.6980v8>
- [23] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014, 15(1): 1929-1958