

基于双向 GRU 神经网络和注意力机制的中文船舶故障关系抽取方法

后同佳 周良

南京航空航天大学计算机科学与技术学院 南京 210016

摘要 随着深度学习的发展,越来越多的深度学习模型被应用到了关系抽取任务中。传统的深度学习模型不能解决长距离的学习任务,且当抽取文本的噪声较大时表现更差。针对以上两个问题,提出了一种基于双向 GRU(Gated Recurrent Unit)神经网络和注意力机制的深度学习模型来对中文船舶故障语料库进行关系抽取。首先,通过使用双向 GRU 神经网络来提取文本特征,解决了文本的长依赖问题,同时减少了模型运行的时间损耗和迭代次数;其次,通过建立句子级别的注意力机制,提高模型对有效语句的关注度,降低噪声句子给整体关系抽取效果带来的负面影响;最后,在训练集上进行训练,并在真实的测试集上计算精确率、召回率、F1 的值来将该模型与现有的方法对比。

关键词: 船舶故障; 门控循环单元; 注意力机制; 关系抽取; 深度学习

中图分类号 TP183

Chinese Ship Fault Relation Extraction Method Based on Bidirectional GRU Neural Network and Attention Mechanism

HOU Tong-jia and ZHOU Liang

School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract With the development of deep learning, more and more deep learning models are applied to relational extraction tasks. The traditional deep learning model can not solve the long-distance learning task, and the performance of the traditional deep learning model is worse when the noise of text extraction is large. To solve the above two problems, a deep learning model based on bidirectional GRU (gated recurrent unit) neural network and attention mechanism is proposed to extract the relationship between Chinese ship faults. Firstly, by using bidirectional GRU neural network to extract text features, the problem of long dependence of text is solved, and the running time loss and iteration times of the model are also reduced. Secondly, by establishing sentence level attention mechanism, the negative impact of noisy sentences on the whole relationship extraction is reduced. Finally, the model is trained on the training set, and the accuracy, recall, and F1 values are calculated on the real test set to compare the model with existing methods.

Keywords Ship fault, Gating cycle unit, Attention mechanism, Relation extraction, Deep learning

1 引言

关系抽取^[1]不仅仅是信息抽取的任务之一,也是构建和补全知识图谱的关键步骤。其主要的研究内容是从文本中抽取实体与实体之间相应的关系,从文本中抽取实体关系也是自然语言处理(Natural Language Processing, NLP)的相关任务。关系抽取具体可以描述为:在一段文本 s 中,需要找出实体 $1(e_1)$ 和实体 $2(e_2)$ 之间存在的关系 r , 或者多种关系 $r_1, r_2, r_3 \dots$ 。

近年来,自然语言处理的任务已经涉及到生活的各个方面。为了解决生活中遇到的搜索、问答、推荐等问题,建立了相关的数据库,如 Freebase^[2], DBpedia^[3] 和 YAGO^[4] 等知识库都在自然语言处理任务中得到广泛的应用。这些数据库中的数据都是由三元组构成的,但随着时代的发展,数据呈现暴增趋势,大数据时代已然到来,人们已经不能满足于这些知识库,迫切地需要从无结构化的文本数据中提取出结构化的数据,尤其是在一些非通用的领域。文献[5]主要探讨了在基于核方法的基础上选取不同语料和最短路径的关系,该方法在

关系提取的性能中取得了不错的进展。文献[6]提出了分析树和递归神经网络的方法,但是没有提取文本的特征信息。文献[7]提出了向双向循环神经网络里逐词输入来提取文本特征信息,但是噪声信息占比高。传统的关系抽取方法都是依据模式匹配,例如 Rink^[8]等提出了通过结合语义关系和词汇的方法, Kambhatla^[9]提出了一种基于逻辑回归的方法。传统的方法会导致计算成本增加并产生额外的传播错误,通用性不佳。Liu 等^[10]提出了基于卷积神经网络的方法,将关系抽取应用到了特定的领域,通过使用卷积神经网络(Convolutional Neural Network, CNN)进行关系分类。Zeng 等人将多实例学习和神经网络模型相结合,建立了基于远程监督的关系提取器,虽然效果不能令人满意,但是该方法在关系提取方面也做出了很大的改进。Huang^[11]提出了基于 GRU 和注意力机制的方法来进行实体之间的关系抽取,提出的模型虽然收敛快,但拟合效果一般。

本文的训练语料库是自建的船舶故障语料库,语料库读取采用 json 文档格式。船舶故障语料库中含有大量的专业术语和相关联的词汇,提取实体间语义关系需要模型拥有长

距离的学习能力。双向 GRU 门控循环单元(Bidirection Gated Recurrent Unit, BiGRU)是循环神经网络(Recurrent Neural Network, RNN)的变形,相比长短时记忆网络(Long Short Term Network, LSTM),其网络同时兼具长距离学习能力,结构更简单,收敛速度更快。因此,本文提出了一种 BiGRU 门控循环单元和注意力机制相结合的实体关系抽取网络模型。所提取的关系有故障原因、解决办法和故障现象 3 个部分。通过词嵌入将预处理的语料库进行向量化的表示;将向量化的词输入到双向的 GRU 循环单元中,获得丰富的上下文语义特征,结合注意力机制,最后通过 softmax 函数提取出最优的结果。

2 关系抽取模型 BiGRU_ATT

2.1 相关介绍

2.1.1 字向量

计算机所识别的语言是数字语言,所以在用深度学习进行自然语言处理任务的时候,需要把数据转换成数学的表达形式,即向量的形式,以方便计算机进行处理和计算。大多数的单词嵌入方式都是以单词作为基本单位,而中文存在特殊性,一个中文词语由好几个字组成,每个字的内部结构信息^[12]也不同,所以本文采用 Word2vector 来训练中文的字向量,字向量是用向量来表示一个字而不是表示一个词,然后选取前 4996 个字,构成词典。给定一个句子 $S = \{e_1, e_2, \dots, e_t\}$,使用 Word2vec 将原始句子中的每一个字都转换成向量 e_i 。首先是将原始句子 S 中的每一个词嵌入到矩阵 $W^{\text{word}} \in R^{d \times |V|}$, V 是固定的词汇表, d 表示词嵌入的大小。句子 W 是需要学习的参数, d 是用户选择的参数,通过矩阵向量将字转换成向量 e_i 。

$$e_i = W^{\text{word}} v^j \quad (1)$$

其中, v^j 表示的是大小为 $|V|$ 的一维向量。由此,便可以把中文语句转换为向量:

$$e_s = \{e_1, e_2, \dots, e_t\} \quad (2)$$

2.1.2 Sigmoid 激活函数

本文采用双向 GRU 神经网络和 Sigmoid 激活函数来提取实体关系。Sigmoid 函数负责将神经元的输入映射到输出中,有增加神经网络模型的非线性变化的作用。该函数是将取值为正负无穷的数映射到 0 和 1 之间,如式(3)所示:

$$g(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

由 Sigmoid 函数的求导推导可知,随着 z 值越来越大或者越来越小,会导致倒数逐渐趋近于 0,从而产生权重消失的现象。推导过程如式(4)所示:

$$g'(z) = \left(\frac{1}{1 + e^{-z}}\right)' = \frac{e^{-z}}{(1 + e^{-z})^2} = g(z)(1 - g(z)) \quad (4)$$

2.2 BiGRU_ATT 模型设计

2.2.1 双向 GRU 网络

GRU 网络是循环神经网络的变形,循环神经网络模型被用来处理时序信息。传统的中文处理模型(Chinese Language Model, CLM)没有考虑到句子之间的上下关联性,而循环神经网络做出的改进是认为与第 n 个字相关联的是前 $n-1$ 个字,循环神经网络可以很好地考虑到长文本的上下文关联问

题。随着隐藏层的增加,普通神经网络会出现梯度消失的问题,因此, Hochreiter 等^[13]提出了长短时记忆模型。LSTM 神经网络模型包括 3 个门:输入门、遗忘门、输出门。输入门用于决定当前时刻有多少输入保留到当前时刻的输入状态;遗忘门决定保留多少上一时刻的状态到当前状态;输出门决定输出多少当前时刻的状态。GRU 网络是 LSTM 网络的简化版本,它仅含有两个门:更新门和重置门。更新门决定前一时刻的信息被输入到当前时刻的量;重置门决定忽略的前一时刻的量。更新门的值越大,说明前一时刻信息被带入的越多;重置门的值越小,说明前一时刻被忽略的信息越多。GRU 网络相比较于 LSTM 网络,网络结构更加简单,不容易产生过拟合现象,收敛和需要迭代的次数也少,且兼具 LSTM 网络模型的功能。GRU 网络结构的前向传播公式如下:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (5)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (6)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \quad (7)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (8)$$

$$y_t = \sigma(W_o \cdot h_t) \quad (9)$$

其中, x_t 表示当前时刻的输入状态;大写字母均为相应的矩阵权重; r_t 表示重置门; z_t 表示更新门; $[]$ 则表示矩阵的连接。循环神经网络的弊端是具有遗忘性,随着输入文本长度的不断增加,信息是有损的,并且输入序列越靠前,信息损失得越严重。本文训练的船舶故障又高度依赖于上下文信息,于是本文采用双向 GRU 神经网络,一个正向学习,一个反向学习;将正向和反向学习的向量连接后作为最后的结构,可以有效避免上下文缺失的问题,双向 GRU 可以有效利用过去和未来的信息。本文所使用的双向 GRU 如图 1 所示,包括前向传播和后向传播两个过程。

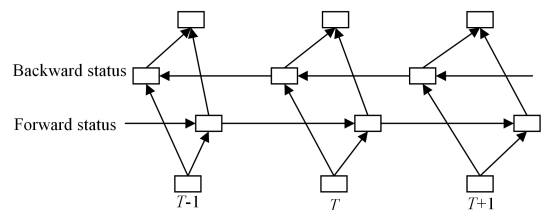


图 1 双向 GRU 神经网络模型

Fig. 1 Bidirectional GRU neural network model

2.2.2 注意力机制

注意力机制的登场源自于在 seq2seq 方法中,解码器能够从编码器中接收到的唯一信息,即最后编码器的隐藏状态能够充分地概括输入序列,然而随着句子长度的增加,其效果很不理想。注意力机制最早被应用在图像领域,Google 的 DeepMind 团队^[14]将注意力机制用于大量的图片识别任务中。实验证明,加入了注意力机制的循环网络比普通的循环网络效果要好很多。随着神经网络的发展,注意力机制也被应用到了卷积神经网络中,并且取得了不错的进展。

传统的方法认为,在无标签的语料库中提取出所有包含实体的句子一定可以表达实体之间的关系,然后这种方法会使提取的句子包含大量的噪音语句,给实体关系的抽取带来很多不好的影响。因此,本文引入了注意力机制,对不同的语料赋予不同的权重,从而能够摒弃一些噪音句

子,提高关系提取的性能。

本文提出了在双向 GRU 神经网络的基础上同时应用注意力机制来提升关系提取的效果。假设双向 GRU 的输出结果组成的矩阵是 $\{h_1, h_2, \dots, h_r\}$, 其中 r 表示句子的长度。句子的表示由输入向量加权形成,公式如下所示:

$$M = \tanh(H) \quad (10)$$

$$\alpha = \text{softmax}(w^T M) \quad (11)$$

$$r = H^{\alpha^T} \quad (12)$$

其中, $H \in R^{d_{in} \times T}$, 是双向 GRU 网络的输入, T 是序列的长度, H 通过 \tanh 函数激活得到 M , 最后 H 乘权重得到输出 r , 再通过 Sigmoid 来得到相应的预测标签的输出。

2.2.3 BiGRU_ATT 模型

关系提取需要高度依赖上下文,因此对网络模型的要求也是需要模型能够及时获取到上下文的信息,在此基础上进行关系的抽取。船舶故障关系的提取也高度依赖上下文信息,需要模型能够正向和反向获取信息,并对信息加以整合。常见的模型忽略对上下文信息的提取,随着 LSTM 网络、GRU 网络和注意力机制的出现,关系提取模型的精度越来越高。BiGRU_ATT 模型利用了双向 GRU 网络的正向和反向传播的优点,迭代时间短;同时,注意力机制可以有效提高网络模型的关注度。

BiGRU_ATT 模型是在双向 GRU 神经网络的基础上加入了注意力机制来提取中文语料库中关于船舶故障的实体关系,总体的网络模型结构如图 2 所示。

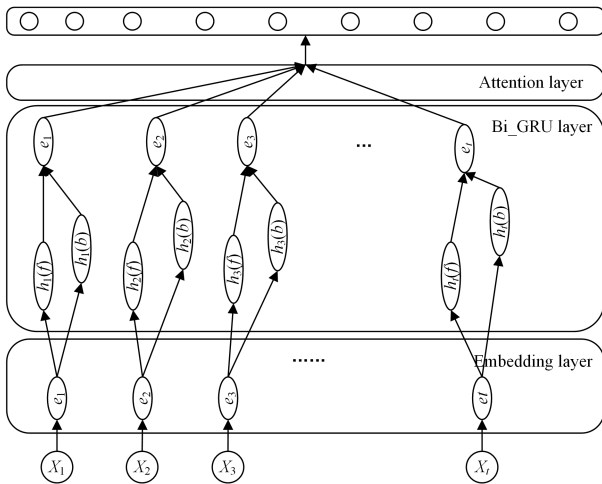


图 2 基于注意力机制的双向 GRU 模型

Fig. 2 Bidirectional GRU model based on attention mechanism

首先,使用字向量构建词典,本文选取了构建词典的前 4996 个字;使用双向 GRU 神经网络来嵌入句子的语义,以此来对句子进行编码,输入的最大长度是 128,构建文本 token、实体 label、关系 label;然后,为了消除噪音数据带来的影响,在特征向量上分别构建注意力机制;最后,提取出权重较高的向量,通过双向 GRU 和 softmax 分类器来提取实体,通过双向 GRU 和 Sigmoid 来提取实体间的关系。由图 2 可知,本文提出的模型共有 5 层:第一层是输入层,输入的是自建的船舶故障语料库中的数据;第二层是嵌入层,将输入数据进行字向量的嵌入;第三层是双向 GRU 层,通过双向 GRU 网络来对句子进行编码;第四层是添加注意力机制,该层是为了降低噪音数据带来的影响,在特征向量上建立句子级别的注意力机制;

最后一层是输出层,通过双向 GRU 神经网络和 Sigmoid 输出关系类型。

2.2.4 数据预处理

(1) 船舶故障语料库

本文所训练的语料库为自建的中文船舶故障语料库,数据来源是各类船舶故障维修手册、船舶故障的书籍和相关的论文、文章。数据集来源有《船舶柴油机检修》《船舶电气检修》《船舶电气设备维护及检测》《船舶辅机检修》《船舶机械修理工艺》《船舶液压设备原理及维修技术》《船舶用螺旋桨原理及维修》《轮机维护与维修》等关于船舶故障的书籍。通过书籍中的相关内容建立原始语料库;之后进行脱敏处理,手工去除部分无用数据和干扰数据。

(2) 人工标注数据

本文训练的数据库是 json 文件格式,为了让模型有较好的提取效果,需要人工标注一部分数据进行模型训练。本文所提取的关系总共有 3 类,分别是故障原因、故障现象和解决措施。人工标注的实例如图 3 所示。

```
{
  text: 阀盘与阀座有严重腐蚀或裂缝时应予以换新
  spo_list: [
    [
      阀盘与阀座有严重腐蚀或裂缝
      解决措施
      应予以换新
    ]
  ]
}
```

图 3 部分语料库人工标注实例

Fig. 3 Examples of manual tagging in some corpora

BiGRU_ATT 模型的原理和构建过程如上文所述,结合文中模型的构建过程,算法 1 给出模型训练的伪代码。

算法 1 BiGRU_ATT 模型训练过程

输入:自建的船舶故障语料库训练语句

输出:故障语料库中关系分类结果

1. train_data=json. data(open(train_test. json))

2. dev_data=json. data(open(test_new. json))

3. classdata_loader().

“加载数据”

4. class Ner_model(tf. keras. Model):

self. char_embedding = tf. keras. layers. Embedding(4996, 64, mask_zero=True)

self. bi_gru = tf. keras. layers. Bidirectional(tf. keras. layers. GRU(6, return_sequence=True))

self. dense_1 = tf. keras. layers. Dense(1)

self. dense_1 = tf. keras. layers. Dense(1)

5. class ER_model(tf. keras. Model):

6. class Extra_result(object):

7. class Evaluate(object):

8. class Attention(Layer):

def __init__():

def build(self, input_shape):

def compute_mask(self, input, input_mask=None):

def call(self, x, mask=None):

def compute_output_shape(self, input_shape):

```

9. model_Ner=Ner_model()
10. model_Er=ER_model()
11. attention=Attention(step_dim=100)
12. optimizer=tf.keras.optimizers.Adam(learning_rate=lr)
13. checkpoint=tf.train.Checkpoint(optimizer=optimizer,model_
    Ner=model_Ner,model_Er=model_Er)
14. evaluate=Evaluate()
15. data_loader=data_loader()
16. best=0.0
17. for epoch in range:
    print('Epoch:',epoch+1)
    print("batch %d:loss %f:loss1 %f:loss2 %f" %)
    print('测试集:', "F %f:P %f:R %f: %" (F,P,F))
18. end

```

3 实验

3.1 数据集和实验设置

目前现有的关系抽取数据库,绝大多数都是英文语料库,针对中文的训练语料库相对较少。本文针对的是中文船舶故障的实体关系抽取,所以采用自建的语料库来进行训练。具体方法是通过搜集各种关于船舶维修的故障手册、船舶故障维修的数据和发表的论文刊物等来获取相应的实体关系语句。本文一共整理了 4987 对确定关系的实体对,共分为 3 类实体关系,分别是故障原因、故障现象、解决措施。针对语料库,本文利用二八规律随机将训练集分为 8:2 后作为训练集和测试集。数据集的具体分类如表 1 所列。

表 1 实体间的各类关系

Table 1 Relationships between entities

实体关系	数量
故障原因	2479
故障现象	436
解决措施	2072

本文所采取的模型的基本参数设置如表 2 所列。

表 2 模型中的参数设置

Table 2 Parameter setting in model

参数	含义	值
num_class	实体关系类别	3
lr	学习率	0.001
num_epoch	迭代次数	22
batch_size	每批次样本数	13
step_dim	Attention 机制的自定义	35

3.2 实验方法和评测指标

为了对比验证本文的实验方法在中文船舶故障的关系抽取效果,分别在双向 GRU 网络、双向 GRU 网络+注意力机制、BiLSTM 网络、BiLSTM 网络+注意力机制网络上对同一数据集进行训练。本文的关系抽取是在特定的领域进行的,可以看做是分类问题,因此对于模型效果,采用了准确率(P)、召回率(R)和 F 值(F)来作为评价指标,计算公式如下:

$$P = \frac{\sum_{i=1}^n P_i}{n} = \frac{\sum_{i=1}^n \text{正确预测为 } i \text{ 类的数目}}{\sum_{i=1}^n \text{预测为 } i \text{ 类的数目}} \quad (13)$$

$$R = \frac{\sum_{i=1}^n R_i}{n} = \frac{\sum_{i=1}^n \text{正确预测为 } i \text{ 类的数目}}{\sum_{i=1}^n \text{属于 } i \text{ 类的实际数目}} \quad (14)$$

$$F = \frac{2 \times P \times R}{P + R} \quad (15)$$

3.3 实验结果和分析

本实验采取的硬件环境是 Intel Core i7, 8 GB 内存, 500 GB 硬盘;软件环境为 Microsoft Windows 10, python3.7。为了验证本文提出的模型的效果,将对自建的故障语料库的同一数据集进行了训练和测试。

(1) 各模型召回率的变化

为了验证 BiGRU_ATT 模型的有效性,将同一数据集在 4 个不同的模型上进行测试,并将测试结果进行对比。数据集的总数据量和最优表现结果如表 3 所列。图 4 所示是不同模型在同一数据集上召回率变化的结果,可以清晰地看到基于双向 GRU 和注意力机制网络的模型结果明显优于其他的网络模型,且整体呈现稳定的趋势,最高可达 0.57 左右。其他 3 个模型的召回率波动较大,可靠性不高。

表 3 不同模型召回率的对比

Table 3 Comparison of recall of different models

模型	总数据量	召回率对比
BiGRU	4987	0.608
BiGRU-attention	4987	0.571
BiLSTM	4987	0.356
BiLSTM-attention	4987	0.307

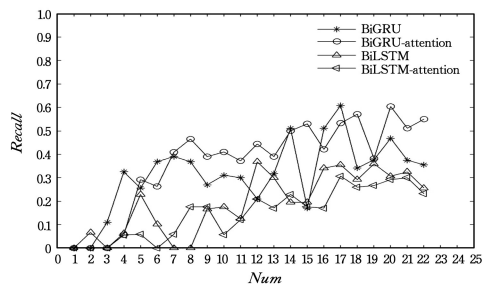


图 4 各模型召回率的对比

Fig. 4 Comparison of recall rate of each model

(2) 各模型精确率的变化

通常将精确率作为模型的评测标准,用精确率的值来衡量模型的部分效果。将同一个自建的船舶故障语料库在 BiGRU_ATT, BiGRU, BiLSTM, BiLSTM_ATT 4 个模型上进行对比测试。表 4 是不同模型的精确率的最高值对比以及总数据量,显然可见 BiGRU_ATT 模型在数据集上的表现相对较好。图 5 给出不同模型上精确率的变化情况,可知基于双向 GRU 和注意力机制的模型的精确率明显优于 BiGRU, BiLSTM, BiLSTM-attention 机制的网络,最高可达 0.923,精确度整体呈现稳定的状态,且整体趋势明显高于其他模型的数值。为了增强可信度,本组实验对同一数据的测试集进行了测试,通过测试集的结果可知, BiGRU_ATT 模型网络结构对中文船舶故障关系的识别可以有效改善提取效果,对多重三元组也有改善。

表 4 不同模型精确率的对比

Table 4 Comparison of precision of different models

模型	总数据量	精确率
BiGRU	4987	0.778
BiGRU-attention	4987	0.928
BiLSTM	4987	0.778
BiLSTM-attention	4987	0.667

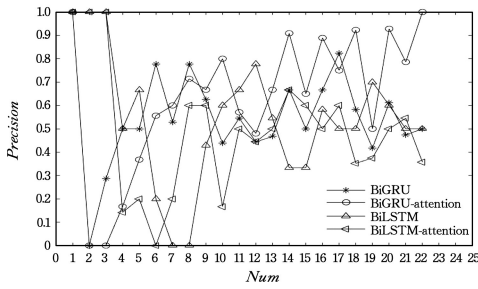


图5 各模型精确率的对比

Fig. 5 Comparison of precision rate of each model

(3)各模型 F1-measure 的变化

表5和图6都分别给出了同一数据集在不同模型上的表现结果。如表5所列,数据集总数据量有4987个,F1的最高值为0.608。从图6可以明显看到,基于双向GRU和注意力机制的网络模型的效果最好,其他3个模型的值波动较大,本文提出的模型相对较稳定,最高值也可以达到0.57左右。

表5 不同模型 F1 结果的对比

Table 5 Comparison of F1-measure of different models

模型	总数据量	F1 值
BiGRU	4987	0.608
BiGRU-attention	4987	0.571
BiLSTM	4987	0.356
BiLSTM-attention	4987	0.307

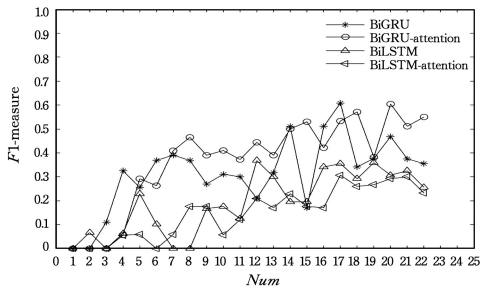


图6 各模型 F1-measure 的对比

Fig. 6 Comparison of F1-measure of each model

结束语 本文提出在基于双向GRU网络的基础上添加注意力机制,用于对中文语料库中船舶故障关系的抽取。该模型采用了字向量的输入,通过双向GRU神经网络对模型进行了建模和特征提取;再通过注意力机制,减少噪音数据的干扰,以此来提高模型的提取效果。通过和双向长短时记忆网络模型以及没有添加注意力机制的模型对比,基于双向GRU网络的注意力机制模型的识别效果更好,且GRU神经单元比LSTM神经单元结构简单,因此训练时间大大减少。实验结果表明本文提出的模型在准确率、召回率等方面都比其他模型的效果更好。由于本文提出的模型是针对船舶故障关系提取这一特定领域,本文的故障语料库为自建的语料库,因此如何继续优化模型的识别效果和增加模型的通用性将是未来继续研究的方向。

参考文献

[1] ZHAO Y Y, QIN B, CHE W X, et al. Research on Chinese event

extraction[J]. Journal of Chinese Information Processing, 2008, 22(1):3-8.

- [2] BOLLACKER K D, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C] // Proc of ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2008: 1247-1250
- [3] AUER S, BIZER C, KOBILAROV G, et al. DBpedia: a nucleus for a Web of open data[C] // Proc of International Semantic Web Conference. Berlin: Springer, 2007: 722-735
- [4] YAGO: a high-quality knowledge base[EB/OL]. <https://www.mpiinf.mpg.de/department/databases-and-information-systems/research/yago-naga/yago/>.
- [5] SUN Z Y, SHUN J Z, YANG J. Chinese entity relation extraction method based on deep learning[J]. Computer Engineering, 2018(9):164-170.
- [6] ZHANG S X. Research on automatic extraction of entity relationship[J]. Journal of Harbin Engineering University, 2006(7):370-373.
- [7] DONG J. Research on feature selection in Chinese entity relation extraction[J]. Chinese Journal of information, 2007(4):80-85.
- [8] RINK B, HARABAGIU S. UTD: classifying semantic relations by combining lexical and semantic resources[C] // Proc of the 5th International Workshop on Semantic Evaluation. Stroudsburg, PA: Association for Computational Linguistics, 2010: 256-259.
- [9] KAMBHATLA N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations [C] // Proc of Interactive Poster and Demonstration Sessions. Stroudsburg, PA: Association for Computational Linguistics, 2004: 22.
- [10] LIU K, FU H D, ZOU Y W, et al. Chinese medical weak supervision relation extraction based on convolution neural network [J]. Computer Science, 2017, 44(10): 249-253.
- [11] HUANG Z W, WEI L, BING C Z, et al. Remote supervised relation extraction based on Gru and attention mechanism [J]. Computer Application Research, 2019(10): 76-79.
- [12] ZHANG S, ZHENG D, HU, et al. Bidirectional long short-term memory networks for relation classification[C] // 29th Pacific Asia Conference on Language, Information and Computation. 2015: 73-78.
- [13] HOCHREITER S, JURGEN S. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [14] MNH V, HEES N. Recurrent Models of Visual Attention [J]. Advances in Neural Information Processing Systems, 2014, 3.



HOU Tong-jia, born in 1997, postgraduate. Her main research interests include information system integration and knowledge graph.