

基于 BERT-GRU-ATT 模型的中文实体关系分类

赵丹丹, 黄德根, 孟佳娜, 董宇, 张攀

引用本文

赵丹丹, 黄德根, 孟佳娜, 董宇, 张攀. [基于 BERT-GRU-ATT 模型的中文实体关系分类](#)[J]. 计算机科学, 2022, 49(6): 319-325.

ZHAO Dan-dan, HUANG De-gen, MENG Jia-na, DONG Yu, ZHANG Pan. [Chinese Entity Relations Classification Based on BERT-GRU-ATT](#)[J]. Computer Science, 2022, 49(6): 319-325.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于混合字词特征的中文短文本分类算法](#)

Chinese Short Text Classification Algorithm Based on Hybrid Features of Characters and Words

计算机科学, 2022, 49(4): 282-287. <https://doi.org/10.11896/jsjx.210200027>

[蒙汉神经机器翻译研究综述](#)

Survey of Mongolian-Chinese Neural Machine Translation

计算机科学, 2022, 49(1): 31-40. <https://doi.org/10.11896/jsjx.210900006>

[融入自注意力机制的深度学习情感分析方法](#)

Self-attention-based BGRU and CNN for Sentiment Analysis

计算机科学, 2022, 49(1): 252-258. <https://doi.org/10.11896/jsjx.210600063>

[基于细粒度差异特征的文本匹配方法](#)

Text Matching Method Based on Fine-grained Difference Features

计算机科学, 2021, 48(8): 60-65. <https://doi.org/10.11896/jsjx.200700008>

[DragDL:一种易用的深度学习模型可视化构建系统](#)

DragDL:An Easy-to-Use Graphical DL Model Construction System

计算机科学, 2021, 48(8): 220-225. <https://doi.org/10.11896/jsjx.200900045>

基于 BERT-GRU-ATT 模型的中文实体关系分类

赵丹丹^{1,2} 黄德根¹ 孟佳娜² 董宇² 张攀²

1 大连理工大学计算机科学与技术学院 辽宁 大连 116024

2 大连民族大学计算机科学与工程学院 辽宁 大连 116600

(zhaodd@dlnu.edu.cn)

摘要 实体关系分类作为自然语言处理的基础任务,对知识图谱、智能问答、语义网构建等任务都起到了非常关键的作用。文中构建了 BERT-GRU-ATT 模型,以进行中文实体关系分类。为消除中文分词歧义对实体关系分类的影响,引入预训练模型 BERT 作为嵌入层,以较好地获得汉字的上下文信息;再通过双向门控循环单元捕获实体在句子中的长距离依赖,通过自注意力机制加强对关系分类贡献明显的字的权重,从而获得较好的实体关系分类结果。为了丰富中文实体关系分类语料,将 SemEval2010_Task8 英文实体关系评测语料翻译为中文¹⁾,该模型在此翻译语料上取得了 75.46% 的 F1 值,说明了所提模型的有效性。此外,所提模型在 SemEval2010-task8 英文数据集上 F1 值达到了 80.55%,证明该模型对英文语料具有一定的泛化能力。

关键词: 中文实体关系分类;预训练模型;门控循环单元;自注意力机制

中图法分类号 TP391

Chinese Entity Relations Classification Based on BERT-GRU-ATT

ZHAO Dan-dan^{1,2}, HUANG De-gen¹, MENG Jia-na², DONG Yu² and ZHANG Pan²

1 School of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024, China

2 School of Computer Science and Engineering, Dalian Minzu University, Dalian, Liaoning 116600, China

Abstract As the basic task of natural language processing, entity relations classification plays a critical role in tasks such as knowledge graphs, intelligent question answering, semantic web construction and so on. This paper constructs the BERT-GRU-ATT model to classify Chinese entity relations. In order to eliminate the influence of Chinese word segmentation ambiguity on entity relations classification, the pre-training model BERT(bi-directional encoder representations from transformers) is introduced as the embedding layer to better obtain the context information of Chinese characters. Then gate recurrent unit(GRU) is used to capture the long-distance dependence of entities in sentences and self-attention mechanism(ATT) is used to strengthen the weight of characters that contribute significantly to relations classification, so as to obtain better results of entity relations classification. In order to enlarge the Chinese entity relations classification corpus, we translate the SemEval2010_Task8 English entity relations evaluation corpus into Chinese. The model achieves an F1 value of 75.46% on this translation corpus, which shows the effectiveness of the proposed model. In addition, the model achieves an F1 of 80.55% on the SemEval2010-Task8 English dataset, which proves that the model has certain generalization ability to English corpus.

Keywords Chinese entity relations classification, Pre-training model, Gate recurrent unit, Self-attention mechanism

1 引言

实体关系抽取工作是自然语言理解和语义分析的核心任务之一,其目标是发现和识别隐含在实体与实体之间的关系。其定义如下:给定一个句子 S , 其中包含实体 $1(E_1)$ 和实体 $2(E_2)$, 预测两个实体间的关系 $r \in R$ (R 是一组关系的集合);或者说,实体关系抽取任务是将文本中无结构化的信息转化

为关系三元组 $\langle E_1, r, E_2 \rangle$ 的形式化描述,其中 E_1 和 E_2 为命名实体, r 为两个实体之间的关系。例如:在句子“俞敏洪创办了新东方”中,抽取关系 $\langle$ 俞敏洪,创始人,新东方 $\rangle$,其中“俞敏洪”和“新东方”被称为实体(Entity),而“创始人”则是他们的关系(Relation)。实体关系抽取研究的成果对信息获取^[1]、自动问答^[2]、机器翻译^[3] 和知识图谱^[4] 等相关领域的研究都有重要价值。

¹⁾ <https://github.com/Chenjiajia123456/Semeval2010Chinese>

到稿日期:2021-06-16 返修日期:2021-10-22

基金项目:国家科技创新 2030-“新一代人工智能”重大项目(2020AAA008004);国家自然科学基金(U1936109,61876031);辽宁省教育厅科研经费(LJYT201906)

This work was supported by the National Key Research and Development Program of China(2020AAA008004), National Natural Science Foundation of China(U1916109,61876031) and Scientific Research Fund of Liaoning Provincial Education Department (LJYT201906).

通信作者:孟佳娜(mengjn@dlnu.edu.cn)

根据给定句子中实体是否已经被明确标注,对该任务做如下分类:若原始文本中已标记出实体,则称该任务为实体关系分类,或者称其为流水线式关系抽取;若输入文本中未标记出句子中的实体,关系抽取任务既要完成实体识别,又要完成关系抽取,则这时该任务被强调为联合关系抽取。后文对关系抽取和关系分类两词的使用不做严格区分。

关系抽取的方法根据其发展历程分为早期基于规则的方法、中期的机器学习方法、目前应用广泛的基于深度学习的方法,以及未来发展方向的开放域关系抽取方法^[5]。

在基于规则的实体关系抽取方法中,主要采用模板匹配的方法来识别关系类型。Deng 等^[6]发现,将英文关系抽取的模板用于中文时效果很差,他们将词汇和语义匹配技术加入模板匹配技术中,使得中文的实体关系抽取效果得以提升。

机器学习方法有基于特征向量和基于核函数方法之分。Che 等^[7]、Guo 等^[8]、Gao 等^[9]根据中文的特点,从不同角度加入中文特征,取得了比基于规则的方法更高的准确性。基于核函数的抽取方法在特征抽取的基础上,根据语法树等解析结构,合理选择核函数,再充分融合各类特征,进一步提高了关系抽取任务的精度。Liu 等^[10]、Guo 等^[11]、Yu 等^[12]利用该方法对中文关系抽取方面进行了研究,效果相比基于特征向量的方法有所提升。

基于深度学习的关系抽取方法将在第 2 节介绍。开放域的关系抽取则针对关系类型的不确定情况产生。Petroni 等^[13]提出的 CORE 模型、Qiu 等^[14]提出的 ZORE 关系抽取模型、Qin 等^[15]完成的 UnCORE 系统都进行了中文开放域二元关系抽取。

2018 年,Google 公司发布预训练模型 BERT (Bi-directional Encoder Representations from Transformers),该模型对自然语言处理中的多项任务效果均有提升。本文将 BERT 应用于中文关系抽取任务,利用 BERT 获取中文文本中汉字的向量表示,从而降低中文文本连续字串造成的分词界限不明确对实体关系抽取造成的影响。该模型首先利用 BERT 作为嵌入层获得中文文本中汉字的上下文语义表示,然后将其送入双向门控单元(Gate recurrent unit,GRU)神经网络以捕获句中词语间的语义关系,再经自注意力机制(Self attention mechanism,ATT)强化对确定实体关系贡献较大的字的权重,最终通过 Softmax 函数输出实体关系类型。本文提出的基于 BERT-GRU-ATT 的中文实体关系分类方法在 SemEval2010_Task8 中文翻译语料上取得了较好的效果。本文的贡献如下:1)建立了 SemEval2010_Task8 中文翻译语料;2)提出了 BERT-GRU-ATT 中文实体关系分类模型,实验结果表明,该模型获得了较好的 F1 值,说明了模型的有效性。

2 相关工作

实体关系分类是判断文本中两个实体之间的关系类型,是知识抽取的核心任务之一。在已有工作的基础上,基于深度学习的有监督关系抽取方法可分为联合学习方法和流水线方法^[4]。本文提出的方法属于基于实体正确标记的流水线方法,该方法通常基于卷积神经网络(Convolutional Neural Network,CNN)、循环神经网络(Recurrent Neural Network,RNN)和长短时记忆(Long and Short Term Memory,LSTM)

神经网络展开。

2.1 基于 CNN 的方法

基于 CNN 的关系分类方法最先由 Zeng 等^[16]于 2014 年提出。该方法中将单词特征和位置特征组合作为输入,通过 CNN 获得句子级特征,再将句子级特征与由两个实体、实体的上下文、以及 WordNet 上位词组成的词汇级特征串联起来,通过 Softmax 函数来确定实体对间的关系。该方法中词级和句子级的特征获得都不需要复杂的预处理,避免了因利用其他 NLP 工具获得词特征而引起的误差传播。

在 Zeng 等解决关系分类方法的基础上,Xu 等^[17]在 CNN 基础上加入了最短依存路径特征,证明了最短依存路径对关系分类的作用。Santos 等^[18]提出的 CR-CNN 模型在 Zeng 等的模型的基础上进行了改进,用排序来取代 Softmax 输出,分类效果有所提升。Ye 等^[19]、Fan 等^[20]、Gao 等^[21]、Sun 等^[22]都采用 CNN 对关系抽取的中文任务进行了尝试。值得注意的是,Sun 等的工作是在 COAE2016 数据集上展开的,由于数据集条目较少,训练集和测试集都不到 1 000 条,因此使用 CNN 模型的效果比传统机器学习方法支持向量机(Support Vector Machine,SVM)模型的 F1 值要低约 10%。由此可以看出,要有足够数量的语料,深度学习模型才能充分发挥作用。

2.2 基于 RNN 的方法

2012 年,Socher 等^[23]提出了循环矩阵-向量模型,该方法为句法分析树的每个节点设置一个向量和一个矩阵,向量代表该节点本身的含义,矩阵则用于表示该成分对上下文的影响。该模型在包括实体关系抽取在内的 3 个自然语言处理子任务上都获得了当时的最好表现。

另外,Lin 等^[24]在 RNN 的基础上引入了注意力机制,Chen 等^[25]提出的基于 RNN 的框架具有较好的扩展性和移植性。

2.3 基于 LSTM 及混合的方法

2015 年,Xu 等^[26]最早使用 LSTM 进行关系抽取。Vu 等^[27]在 Santos 等^[18]提出的模型的基础上进一步将 CNN 与双向 RNN 相结合来提取关系。语境分为左、中、右 3 个不相交的语境区域,这种方法重复利用两个实体之间的中间语境,最后采用一种投票机制将 CNN 和 RNN 模型结果有机结合,得到了当时的最好效果。

Schlichtkrull 等^[28]和 Zhang 等^[29]首先将图神经网络用于实体关系抽取,但效果不理想。Song 等^[30]和 Guo 等^[31]则进一步将图神经网络与 LSTM 及注意力机制相结合,用于进一步提高抽取精度。另外,深度信任网络也被 Chen 等^[32]应用于中文名实体关系抽取,效果较好。

以上深度学习方法在英文语料上应用广泛,大多效果较好,但在中文关系抽取领域,由于中文文本中字与词的关系不同于英文文本的字母与单词界限明晰、表意明确,并且中文一词多义现象普遍,句式表达复杂,这无疑增加了中文关系抽取任务的难度。而且中文实体关系的标注语料很少,这进一步影响了中文实体抽取的研究。

SemEval2010_Task8 语料在关系抽取领域使用广泛,本文通过翻译该语料建立起中文关系抽取数据集,同时使用 BERT 模型来获得与上下文相关的汉字稠密向量,进而采用

双向 GRU 模型获得句子中的长距离依赖关系,并通过自注意力机制捕获句子的深层语义信息,从而得到较好的中文关系分类结果。

3 BERT-GRU-ATT 模型

上述工作中,研究者对实体关系抽取方法进行了广泛的研究和尝试,但实体关系抽取效果对中文文本还有较大的提升空间。本文提出了采用 Google 发布的针对中文文本的预训练 BERT 模型,并结合比较成熟的 GRU 和自注意力机制对中文文本进行实体关系抽取。

BERT-GRU-ATT 模型架构如图 1 所示,该模型包含输入层、BERT 嵌入层、双向 GRU 层、自注意力机制层和 Softmax 输出层。输入层采用 BERT 预训练模型,获得低维稠密向量来表达汉字语义,继而将其输入双向 GRU 层,得到句子级的上下文特征,对双向 GRU 最终的隐层输出再加入自注意力机制,调整单个汉字对实体关系贡献大小的编码权重后,将其输入 Softmax 层,以确定最终的关系类型。各层的操作机制如下:

(1)输入层:模型以实体 1、实体 2 和句子为输入。句子不分词,以单个汉字为输入单位,为确定实体在句子中的位置,句中实体通过“#”号代替。

(2)BERT 层:采用 BERT 预训练模型,将输入的汉字转换为包含字一级上下文信息的低维稠密向量表示。我们采用的是 BERT-Base 中文版。

(3)GRU 层:接收预训练模型的输出结果,通过双向的 GRU 获得句子级的上下文特征。这里采用 GRU 是出于 3 个方面的考虑:首先 GRU 作为 LSTM 模型的变种,解决了 RNN 模型的梯度弥散和梯度爆炸问题;其次 GRU 为句子中的诸汉字提供长距离依赖的支持;最后 GRU 的运算较 LSTM 略快,并且在规模不是很大的语料上效果略好。

(4)ATT 层:采用自注意力机制加强对关系分类贡献明显的字的权重。通过自注意力机制,充分利用句子的内部结构,使句子中确定关系类型的关键信息得以凸显。

(5)输出层:该层接收 ATT 层的输出,通过一个全连接和 Softmax 作用输出句中两实体间特定的关系类型。Softmax 为一个逻辑回归分类器,可一次性完成多分类问题。

3.1 BERT 层

本文采用 Google 发布的 BERT 作为嵌入层,它能有效地将字序列变换为稠密向量。特别地,由于汉语区别于其他语言,因此本文采用专门的汉语 BERT 预训练模型获得汉字的先验知识。预训练的 BERT 利用双向的 transformer 作为编码,使得每个字的表示中合并了字左右方向的信息。具体来说,BERT 模型不是使用语言模型来进行预训练,也不采用双向的操作,而是用两个特殊任务对 BERT 模型进行预训练。这两个任务分别是掩码语言模型(Mask Language Model, MLM)和下一个句子预测(Next Sentence Prediction, NSP)。MLM 是将输入串中的 15% 随机屏蔽,然后通过训练模型对被屏蔽的字串进行预测,这样可以在更深的层次上得到句子的双向表示。NSP 任务旨在学习、理解不同句子之间关系的模型,即输入两句的“isNext”关系将被预测为“yes”或“no”,表示第二个句子是否是第一个句子的下一句。采用 MLM 与

NSP 两个任务相结合的方式,在预训练阶段,大量的语义信息被整合到模型中,相比通过 word2vec 得到的词向量信息更丰富、表意更准确。

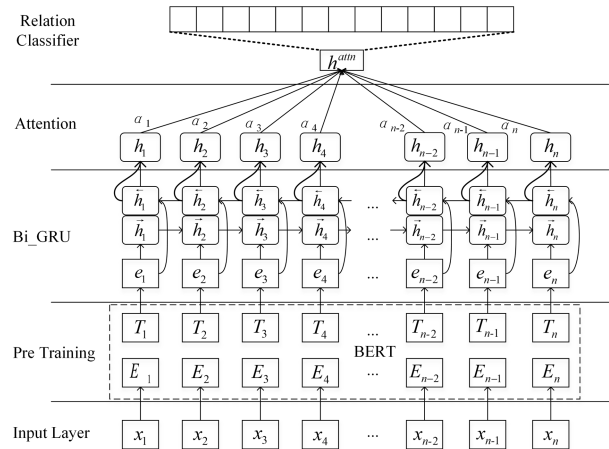


图 1 BERT-GRU-ATT 模型架构

Fig. 1 Architecture of BERT-GRU-ATT model

本文采用预训练的 BERT 模型作为嵌入层,为了提高训练效率,在训练时冻结 BERT 层。图 2 给出了 BERT 的具体使用形式。为了更好地发挥每个句子实例的代表价值,本文采用的输入方案是在句子开始部分就将句中实体 1 和实体 2 输入模型,而在句子中实体出现的部分以“#”替换。通过这种方式能让模型更好地学习到该句式两实体间的关系,而不局限于特定的实体。同时,前序研究表明句子中每个字与两个实体的距离对实体关系识别很有价值,句子中以“#”替代实际实体的输入方案,不需要再额外标注距离便可有效地传递出该信息。

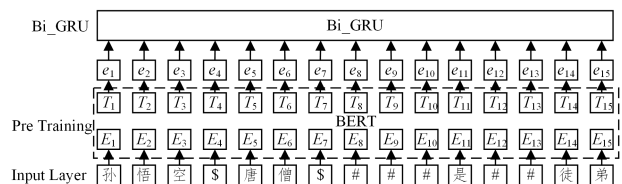


图 2 BERT 层句子输入模式

Fig. 2 Input sentence mode in BERT layer

设给定的输入汉字为 x_i , BERT 嵌入层的输出表示为 e_i , 则其变换如式(1)所示:

$$e_i = \text{BERT}(x_i) \quad (1)$$

3.2 GRU 层

为了解决循环神经网络 RNN 的梯度弥散问题, LSTM 和 GRU 模型被广泛采用。LSTM 核心结构涉及 3 个门,即遗忘门、输入门和输出门,计算量相对较大。而 GRU 模型只有更新门和复位门两个门,即图 3 中的“ z ”门和“ r ”门。更新门控制前一时刻的状态信息进入当前状态的程度,更新门的值越大,前一时刻的状态信息被带入得就越多。复位门是为了控制忽略先前状态信息的程度,复位门的值越小,越容易被忽略。

$$r_t = \sigma(W_{er}e_t + W_{hr}h_{t-1} + b_r) \quad (2)$$

$$z_t = \sigma(W_{ez}e_t + W_{hz}h_{t-1} + b_z) \quad (3)$$

$$\tilde{h}_t = \tanh(W_{eh}e_t + W_{hh}(r_t \cdot h_{t-1}) + b_h) \quad (4)$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \cdot \tilde{\mathbf{h}}_t + \mathbf{z}_t \cdot \mathbf{h}_{t-1}$$

(5)

$$\mathbf{h}_t = \vec{\mathbf{h}}_t \oplus \overleftarrow{\mathbf{h}}_t$$

(6)

式(2)–式(6)说明了 GRU 单元的计算过程。设前一步的隐含状态为 $\mathbf{h}_{t-1}$,当前输入为 $\mathbf{e}_t$,对应的偏置参数为 $\mathbf{b}_r$ 、 $\mathbf{b}_z$ 和 $\mathbf{b}_h$ 。复位门和更新门的计算式如式(2)、式(3)所示。 σ 为 sigmoid 激活函数, $\mathbf{W}_{er}$ 、 $\mathbf{W}_{hr}$ 、 $\mathbf{W}_{ez}$ 、 $\mathbf{W}_{hz}$ 、 $\mathbf{W}_{eh}$ 和 $\mathbf{W}_{hh}$ 是对应的权重参数矩阵。候选隐态定义为 $\tilde{\mathbf{h}}_t$,如式(4)所示。隐态的定义如式(5)所示,它是候选隐态 $\tilde{\mathbf{h}}_t$ 与前一时刻隐层输出 $\mathbf{h}_{t-1}$ 的线性组合,符号 $\cdot$ 代表向量对应元素相乘。

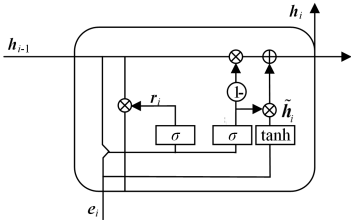


图 3 GRU 单元结构
Fig. 3 Cell structure of GRU

标准 GRU 以从前到后的顺序来处理序列,这样它无法获得未来的信息。双向 GRU 通过将正向和反向 GRU 相结合,来同时获取过去和未来的信息。整个 GRU 层的输出如式(6)所示,其中 $\vec{\mathbf{h}}_t$ 代表前向输出, $\overleftarrow{\mathbf{h}}_t$ 代表后向输出,符号 $\oplus$ 代表将两个向量进行拼接。

3.3 ATT 层

自注意力机制描述为将查询和一组键值对映射到输出,其中查询、键、值和输出都是向量,输出计算为值的加权和,其中分配给每个值的权重由查询与相应的键计算获得,给定双向 GRU 层的输出层 $\mathbf{h}=[\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N]$,采用线性变换获得查询、键和对应的值,如式(7)–式(9)所示:

$$\mathbf{q}_i = \mathbf{W}_q \mathbf{h}_i + \mathbf{b}_q$$

(7)

$$\mathbf{k}_i = \mathbf{W}_k \mathbf{h}_i + \mathbf{b}_k$$

(8)

$$\mathbf{v}_i = \mathbf{W}_v \mathbf{h}_i + \mathbf{b}_v$$

(9)

其中, $\mathbf{q}_i$ 、 $\mathbf{k}_i$ 和 $\mathbf{v}_i$ 分别代表查询向量、键向量和值向量, $\mathbf{W}_q$ 、 $\mathbf{W}_k$ 和 $\mathbf{W}_v$ 分别为对应参数矩阵, $\mathbf{b}_q$ 、 $\mathbf{b}_k$ 和 $\mathbf{b}_v$ 分别为对应的偏置向量。计算 Attention 的过程如图 4 所示。

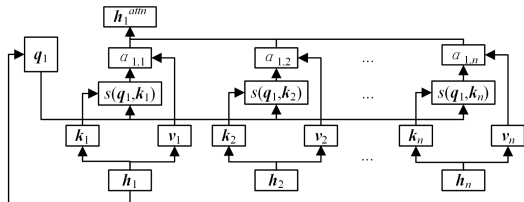


图 4 自注意力机制示意图
Fig. 4 Schematic diagram of self-attention mechanism

利用 $\mathbf{q}_i$ 和 $\mathbf{k}_i$ 计算每两个输入向量之间的相关性,即计算每两个输入向量的得分,采用点乘的方式,计算式如下:

$$s(\mathbf{q}_i, \mathbf{k}_j) = (\mathbf{q}_i)^T \cdot \mathbf{k}_j$$

(10)

对计算的得分进行归一化处理,得到最终的得分 $a_{i,j}$,计算式如下:

$$a_{i,j} = \frac{\exp(s(\mathbf{q}_i, \mathbf{k}_j))}{\sum_{l=1}^N \exp(s(\mathbf{q}_i, \mathbf{k}_l))}$$

(11)

其中, $i=1,2,\dots,N$, N 代表句子中所有的字。最终的输出向量的计算式如下:

$$\mathbf{h}_i^{attn} = \sum_{j=1}^N a_{i,j} \mathbf{v}_j$$

(12)

4 实验

4.1 实验语料

SemEval2010_Task8 的关系抽取语料在实体关系抽取研究中应用广泛,其语料包含 9 种关系类型,8000 条训练语料,2717 条测试语料。该语料规模适中,语料中各类关系的语句条数基本均衡,因此本文选择它作为源语料进行对应的中文实体关系抽取语料建设。

源语料总数超过 1 万条,一个人完成翻译除了工作量大以外,翻译质量也无法保证。在该中文实体关系抽取语料的获取过程中,本文采用了 3 级处理:课题组带领 10 个小组,每组 1 个组长,每个组长带领 10 个成员。每个成员完成 100 条语句的翻译后汇总给 10 个小组长;各小组长完成个人翻译任务外,抽样人工校对组员上交的成果,以进一步确保翻译质量;最后课题组再进一步校订各个小组上交的翻译结果。

对比识别不准确语句对中英翻译中遇到的语言点,做出以下总结。

(1) 语序问题

由于英文中以中心词在前,修饰语在后,而汉语则惯常于中心词在后,修饰语在前,因此在某些情况下,两个实体会因翻译而变换顺序,如:

原文: A small ceramic $\langle e1 \rangle$ figurine $\langle /e1 \rangle$ of a baby was hidden in the $\langle e2 \rangle$ cake $\langle /e2 \rangle$ 。

译文: $\langle e2 \rangle$ 蛋糕 $\langle /e2 \rangle$ 里藏着一个 $\langle e1 \rangle$ 小陶瓷 $\langle e1 \rangle$ 雕像 $\langle /e1 \rangle$ 。

或 $\langle e2 \rangle$ 蛋糕 $\langle /e2 \rangle$ 里藏着一个 $\langle e1 \rangle$ 瓷娃娃 $\langle /e1 \rangle$ 。

这类句子常常会导致关系方向的识别错误。

(2) 表达问题

对于有些语句,本文采用直译会保留实体间的关系,如下面的语句,葡萄酒(wine)和塑料盒(carton)两实体之间是内容-容器(Content-Container)的关系,但有时我们不按字面意思,将其修改为意译的结果,反而在实验中会无法识别两个实体间的关系。

原文: When I first arrived in Italy I was amazed to find that they sell $\langle e1 \rangle$ wine $\langle /e1 \rangle$ in a $\langle e2 \rangle$ carton $\langle /e2 \rangle$ 。

直译: 当我刚到意大利时,我惊奇地发现他们把 $\langle e1 \rangle$ 葡萄酒 $\langle /e1 \rangle$ 放到 $\langle e2 \rangle$ 塑料盒 $\langle /e2 \rangle$ 里来卖。

意译: 当我刚到意大利时,我惊奇地发现他们是按 $\langle e2 \rangle$ 盒 $\langle /e2 \rangle$ 卖 $\langle e1 \rangle$ 葡萄酒 $\langle /e1 \rangle$ 的。

(3) 一词多义问题

英文中也有很多多义词,我们常常按最常用的词义来翻译语句,但这样的理解未必全部正确。

原文: The United States has $\langle e1 \rangle$ laws $\langle /e1 \rangle$ that limit magazine $\langle e2 \rangle$ size $\langle /e2 \rangle$, too.

错误译文: 美国也有限制杂志 $\langle e2 \rangle$ 大小 $\langle /e2 \rangle$ 的 $\langle e1 \rangle$ 法律 $\langle /e1 \rangle$ 。

正确译文: 美国也有限制杂志 $\langle e2 \rangle$ 规模 $\langle /e2 \rangle$ 的 $\langle e1 \rangle$ 法律 $\langle /e1 \rangle$ 。

还有的句子经过翻译以后,原句中的实体变得模糊或消失,经过必要的调整,我们得到了 SemEval2010_Task8 的中文翻译语料。

4.2 实验设置

本文采用 Google 发布的中文 BERT 预训练模型展开实验。为加速方法的训练和验证,训练模型时,BERT 的参数被锁定以减小预训练语言模型的影响。实验设置如表 1 所列。

表 1 实验环境参数

Table 1 Experiment environment parameters

操作系统	Windows10
GPU	Nvidia GeForce RTX 2080 Ti
CPU	Intel(R) Core i9-9900K
CUDA	10.2.89
CUDNN	7.6.5
Python	3.6.10
Tensorflow	1.13.0

在该中文实体关系分类任务中,BERT 维度为 768,双向 GRU 隐层状态是 128,自注意力机制权重矩阵为 32。各参数矩阵及偏置向量随机初始化。

4.3 性能评价方案

性能评价采用该领域常用的精确率(P)、召回率(R)和 F1 值,如式(13)一式(15)。由于实体关系抽取属于多分类问题,因此总的性能评价采用的是微平均,即每个实例性能指标的算数平均。

$$P=\frac{TP}{TP+FP}\times100\%$$
 (13)

$$R=\frac{TP}{TP+FN}\times100\%$$
 (14)

$$F1=\frac{2RP}{P+R}\times100\%$$
 (15)

其中,TP 为正确判断为该类别的实例数;FP 为判断为该类别,但实际并非属于该类的实例数;FN 表示属于该类别,但未被判定为该类别的实例数。

4.4 实验结果及对比分析

本文在 SemEval-2010_Task8 翻译语料上评估了本文模型,该语料包含 10717 个标注好的示例,其中 8000 个用于训练,2717 个用于测试。本文从训练数据中随机抽取了 2182 个样本进行验证。

4.4.1 基础实验

GRU 是较为常见的长短时记忆网络的一种变体,而 LSTM 也是以往关系抽取任务中的主流模型。为了对比分析分别使用 GRU 和 LSTM 时对模型性能的影响,将 BERT-GRU-ATT 模型中的 GRU 替换为 LSTM 进行了实验,实验结果如表 2 所列。

表 2 GRU 与 LSTM 模型的对比结果

Table 2 Comparison results of GRU and LSTM models

(单位:%)

模型	P	R	F1 值
BERT+LSTM	74.38	74.5	74.44
BERT+GRU	75.42	73.94	74.67
BERT+LSTM+ATT	74.71	75.46	75.08
BERT+GRU+ATT	75.55	75.39	75.46

表 2 中,BERT+LSTM 和 BERT+GRU 分别为未加入 ATT 的模型,BERT+LSTM+ATT 为本文模型的变体

模型。实验中双向 GRU 模型的实验结果都高于对应的双向 LSTM 模型,原因是该翻译的语料训练加测试数据共有 10717 条,语料集规模较小。同时,相对于 LSTM 模型,GRU 简单,训练时所需参数更少,并不需要依赖于大量的语料。因此,在较小规模语料集上,双向 GRU 模型可以取得更好的效果。此外,在大规模语料集上,即便 LSTM 的表达性能更好,同样因为 GRU 比 LSTM 模型需要训练的参数更少,更容易收敛,能节省很多时间,其效率比 LSTM 更高。

4.4.2 分类别实验

为了确认模型对不同关系分类效果的差别,分别对数据集中的各个类别进行了精确率(P)、召回率(R)和 F1 值的计算,结果如表 3 所列。

表 3 分类别实验的结果

Table 3 Results of classification experiments

关系类别	关系类型	F1 值/%
Cause-Effect	原因-结果	82.39
Component-Whole	部分-整体	69.35
Content-Container	内容-容器	80.10
Entity-Destination	实体-目的地	86.67
Entity-Origin	实体-来源	75.10
Instrument-Agency	工具-使用者	64.83
Member-Collection	成员-集合	72.77
Message-Topic	消息-主题	76.34
Product-Producer	产品-生产商	71.82

从表 3 中可以看出,除无关类外,实体-目的地(Entity-Destination)类别的识别效果最高,F1 值达到了 86.67%,而工具-使用者(Instrumnet-Agency)关系识别结果最低,F1 值为 64.83%。对工具-使用者关系中识别错误的语句进行统计,未能识别出两实体间有关系的语料有 39 条,除 14 条为关系识别错误,有 25 条为方向识别错误。可见识别实体关系的方向是一个难点。

4.4.3 分方向实验

对整个实验结果进行统计,在 541 条错误预测结果中有 237 条为未能判断出句子中两个实体存在关系,67 条为方向预测错误。可见关系的方向识别对实验结果有很大影响,因此本文进行了不区分关系方向的对比实验。实验结果如表 4 所列。

表 4 区分关系方向与否的实验对比

Table 4 Experimental comparison of distinguishing direction of relationship or not

(单位:%)

	区分方向	不分方向
P	75.55	78.38
R	75.39	78.22
F1	75.46	78.23

由实验结果可以看出,不区分方向的结果优于区分方向的结果 3%左右。

不分方向抽取效果更好的原因分析如下:首先,不分方向关系类别减少了一半,关系类别数的降低对结果的影响比较直接;其次,在同一关系中判断关系的方向相对精细,难度系数高于不同类别间的区分,因此识别结果差距明显。

4.4.4 消融实验

为验证不同模块对模型总体性能的影响,本文进行了消融实验,实验结果如表 5 所列。

表5 模型不同变体的比较

Table 5 Comparison of model variations

模型组合	(单位: %)		
	P	R	F1 值
GRU+ATT(Baseline)	61.08	57.46	59.21
BERT	61.94	60.33	60.94
BERT+ATT	69.20	69.47	69.33
BERT+GRU	75.42	73.94	74.67
BERT+GRU+ATT(本文)	75.55	75.39	75.46

采用双向 GRU 和自注意力机制的模型(GRU+ATT)为 Baseline。该模型没有 BERT 层,而是使用 word2vec 词向量和句子中的词分别与两个命名实体的距离为输入。BERT 模型是直接将 BERT 输入进行全连接后采用 Softmax 输出分类结果。BERT+ATT 代表将 BERT 输出经自注意力机制作用后直接分类的模型。BERT+GRU 代表 BERT 输出送入双向 GRU 输入,不经自注意力机制处理的分类模型。BERT+GRU+ATT 则为本文提出的 BERT 加双向 GRU 再经注意力机制作用后分类输出的模型。

通过对模型各个不同变体的比较发现,仅单独的 BERT 模型,就与作为 Baseline 的 GRU+ATT 模型程度相当,且略有高出。因为 BERT 模型作为大规模预训练模型,自身的编码层有自注意力机制,同时模型又可以捕捉到较为长距离的信息,从而提升了预测准确率。在 BERT 模型的基础上加入 ATT 或者 GRU 后,效果都有提升,但 BERT+GRU 比 BERT+ATT 的效果高出 5.34%,因为相比 ATT,GRU 更易获取长距离特征信息,可以充分利用上下文信息,说明长距离的上下文特征信息对于模型最终的分类效果有很大的帮助。进一步地,BERT+GRU+ATT 模型比 BERT+GRU 模型的效果又提升了 0.79%,说明注意力机制能充分获取到上下文特征信息内部的有用特征,进一步提高了模型的特征表达能力,同时也较大程度地排除了冗余信息的干扰,因此取得了最优的效果。

4.4.5 泛化实验

为防止过拟合,基础实验在 GRU 模块直接初始化丢失率(dropout)的值为 0.5。为了进一步探讨 dropout 对实验结果的影响,将 dropout 分别设置为 0.1~0.9,得出实验曲线如图 5 所示。

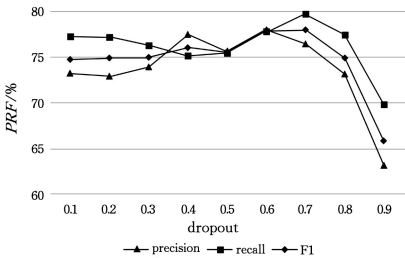


图5 Dropout 对实验结果的影响

Fig. 5 Influence of dropout on experimental results

Dropout 在前向传播时让某些神经元以一定的概率值停止工作,这样可以使模型的泛化性更强,因为这样模型不会太依赖于某些局部特征。该模型在 dropout 为 0.7 时性能达到最优,此时神经元有 70% 的概率被保留下来,30% 的概率被失活,可以很好地学习特征信息的同时,又能够防止过拟合。

将模型的中文 BERT 换为英文 BERT,在 SemEval2010_

Task8 英文语料上进行了实验,与选取 dropout 为 0.7 时的中文模型结果相比,二者的精确率已经相差不大,实验结果如表 6 所列。由于英文模型的召回率较高,总体 F1 值相差不到 3%,此性能差距对中英文语料而言相对较小。

表6 中英文 SemEval2010_Task8 语料上的实验结果对比

Table 6 Comparison of experimental results on SemEval2010_Task8 Chinese and English corpora

	(单位: %)	
	中文	英文
P	76.40	77.41
R	79.61	84.19
F1	77.92	80.55

结束语 关系抽取是自然语言处理领域中的关键任务,本文研究了中文实体关系分类任务,提出了 BERT-GRU-ATT 中文实体关系分类方法:首先采用 BERT 获取中文汉字中的隐含语义向量,将字特征及字的上下文特征都通过 BERT 嵌入层一次性赋予模型,不需要人工干预,有效降低了成本和工作量。本文模型采用双向 GRU 获得句中长距离依赖的句子信息,通过自注意力机制有效提高了句子中对关系判断贡献较大的词的比重,以此进一步提升模型的效果。

本文在 SemEval2010_Task8 语料的基础上完成了翻译工作,并展开了有效的实验。实验结果表明 BERT-GRU-ATT 模型可以有效抽取中文句子中的实体关系。

参考文献

[1] XU J,ZHANG Z X,WU Z X. Review on Techniques of Entity Relation Extraction[J]. Data Analysis and Knowledge Discovery, 2008,24(8):18-23.

[2] LIU Q,LI Y,DUAN H,et al. Knowledge Graph Construction Techniques[J]. Journal of Computer Research and Development,2016,53(3):582-600.

[3] LIU Y J. Research on Automatic Extraction of Chinese Named Entities and Entity Relations[D]. Zhengzhou: Zhengzhou University,2019.

[4] E H H,ZHANG W J,XIAO S Q,et al. Survey of Entity Relationship Extraction Based on Deep Learning[J]. Journal of Software,2019,30(6):1793-1818.

[5] LI D M,ZHANG Y,LI D Y,et al. Review of Entity Relation Extraction Methods [J]. Journal of Computer Research and Development,2020,57(7):1424-1448.

[6] DENG B,FAN X Z,YANG L G. Entity Relation Extraction Method Using Semantic Pattern [J]. Computer Engineering, 2007,33(10):212-214.

[7] CHE W X,LIU T,LI S. Automatic Entity Relation Extraction [J]. Journal of Chinese Information Processing,2005,19(2):2-7.

[8] GUO X Y,HE T T,HU X H,et al. Chinese Named Entity Relation Extraction Based on Syntactic and Semantic Features[J]. Journal of Chinese Information Processing, 2014, 28(6): 183-189.

[9] GAO J P,ZHANG H,ZHAO X J,et al. Evolutionary Relation Extraction for Domain Knowledge in Wikipedia [J]. Chinese Journal of Computers,2016,39(10):2088-2101.

[10] LIU K B,LI F,LIU L,et al. Implementation of a Kernel-Based Chinese Relation Extraction System[J]. Journal of Computer

- Research and Development, 2007, 44(8): 1406-1411.
- [11] GUO J Y, CHEN P, YU Z T, et al. Domain Specific Chinese Semantic Relation Extraction Based on Composite Kernel[J]. Journal of Chinese Information Processing, 2016, 30(1): 24-30.
 - [12] YU H H, QIAN L H, ZHOU G D, et al. Chinese Semantic Relation Extraction Based on Unified Syntactic and Entity Semantic Tree[J]. Journal of Chinese Information Processing, 2010, 24(5): 7-23.
 - [13] PETRONI F, CORRO L, GEMULLA R. Core: Context-aware open relation extraction with factorization machines[C]// Proceedings of the 2015 Conference on Empirical Methodes in Natural Language Processing. Lisbon, Portugal, 2015: 1763-1773.
 - [14] QIU L K, ZANG Y, ZORE. A syntax-based system for Chinese open relation extraction[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2014: 1870-1880.
 - [15] QIN B, LIU A A, LIU T. Unsupervised Chinese Open Entity Relation Extraction[J]. Journal of Computer Research and Development, 2015, 52(5): 1029-1035.
 - [16] ZENG D, LIU K, LAI S, et al. Relation classification via convolutional deep neural network[C]// Proceedings of the 25th International Conference on Computational Linguistics; Technical Papers (COLING 2014). Dublin, Ireland, 2014: 2335-2344.
 - [17] XU K, FENG Y, HUANG S, et al. Semantic relation classification via convolutional neural networks with simple negative sampling[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015: 941-949.
 - [18] SANTOS C N D, XIANG B, ZHOU B. Classifying relations by ranking with convolutional neural networks[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China, 2015: 132-137.
 - [19] YE H, CHAO W H, LUO Z C, et al. Jointly extracting relations with class ties via effective deep ranking[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, 2017: 1810-1820.
 - [20] BAI F, RITTER A. Structured minimally supervised learning for neural relation extraction[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Minneapolis, Minnesota, 2019: 3057-3069.
 - [21] GAO D, PENG D L, LIU C. Entity Relation Extraction Based on CNN in Large-scale Text Data[J]. Journal of Chinese Computer Systems 2018, 39(5): 1021-1026.
 - [22] SUN J D, GU X S, LI Y, et al. Chinese entity relation extraction algorithms based on COAE2016 datasets [J]. Journal of Shandong University(Natural Science) 2017, 52(9): 7-12, 18.
 - [23] SOCHER R, HUVAL B, MANNING C D, et al. Semantic Compositionality through Recursive Mtrix-Vector Spaces[C]// Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Prcessing and Computational Natural Language Learning. Stroudsburg, PA, UA, 2012: 1201-1211.
 - [24] LIN Y K, SHEN S Q, LIU Z Y, et al. Neural relation extraction with selective attention over instances[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016: 2124-2133.
 - [25] CHEN L, TIMOTHY M, DMITRIY D, et al. Self-training improves recurrent neural networks performance for temporal relation extraction[C]// Proceedings of the 9th International Workshop on Health Text Mining and Information Analysis. Brussels, Belgium, 2018: 165-176.
 - [26] XU Y, MOU L L, LI G, et al. Classifying relations via long short term memory networks along shortest dependency paths[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015: 1785-1794.
 - [27] NGOC T V, ADEL H, GUPTA P, et al. Combining recurrent and convolutional neural networks for relation classification [C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics. San Diego, California, 2016: 534-539.
 - [28] SCHLICHTKRULL M S, KIPF T, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// Proceedings of the European semantic Web Conference on. Heraklion, Crete, Greece, 2018: 593-607.
 - [29] ZHANG Y H, QI P, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2205-2215.
 - [30] SONG L F, ZHANG Y, WANG Z G, et al. N-ary relation extraction using graph state LSTM[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018: 2226-2235.
 - [31] GUO Z J, ZHANG Y, LU W. Attention guided graph convolutional networks for relation extraction[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 241-251.
 - [32] CHEN Y, ZHENG D Q, ZHAO T J. Chinese relation extraction based on Deep Belief Nets [J]. Journal of Software, 2012, 23(10): 2572-2585.



ZHAO Dan-dan, born in 1975, postgraduate, associate professor, is a member of China Computer Federation. Her main research interests include nature language processing and information extraction.



MENG Jia-na, born in 1972, Ph.D, professor, is a member of China Computer Federation. Her main research interests include nature language processing, transfer learning and sentiment analysis.