

文章编号: 1003-0077(2020)11-0096-08

## 基于全局和局部特征感知网络的关系提取方法

宋 威<sup>1,2</sup>, 朱富鑫<sup>1</sup>

(1. 江南大学 人工智能与计算机学院, 江苏 无锡 214122;

2. 江南大学 江苏省模式识别与计算智能工程实验室, 江苏 无锡 214122)

**摘 要:** 关系提取旨在从大量的非结构化数据中提取实体对间的关系。目前, 大多数神经网络仅考虑单分支信息流, 难以获取足够的语义特征来进行关系提取。针对此问题, 该文提出一种基于全局和局部特征感知网络的关系提取方法。该方法首先利用自注意力机制和循环神经网络获取单词的相关性时序特征。其次, 构建多分支特征感知卷积神经网络, 分别获取全局和局部相关性时序特征, 避免全局和局部感知的相互影响。进一步地, 对这两种特征进行拼接和筛选, 以全面表示句子的重要语义特征。最后, 结合 Softmax 分类器实现关系提取。实验结果表明, 该文所提方法优于主流的基于卷积神经网络和循环神经网络的关系提取方法, 在标准的 SemEval-2010 Task 8 和 KBP37 数据集上  $F_1$  值分别达到 86.1% 和 64.9%。

**关键词:** 关系提取; 神经网络; 自注意力机制; 特征感知; 多分支

**中图分类号:** TP391

**文献标识码:** A

## Global and Local Feature-Aware Network for Relation Extraction

SONG Wei<sup>1,2</sup>, ZHU Fuxin<sup>1</sup>

(1. School of Artificial Intelligence and Computer Science, Jiangnan University,

Wuxi, Jiangsu 214122, China;

2. Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi, Jiangsu 214122, China)

**Abstract:** Relation extraction aims to identify the relationship between entities from a large amount of unstructured data. This paper proposes a relation extraction method based on global and local feature-aware network. This method first adopts the self-attention mechanism and recurrent neural network to obtain the correlated sequence features of each word. Then, a multi-branch feature-aware convolutional neural network is constructed to obtain global and local features without their mutual interference. Moreover, the obtained two features are concatenated to fully represent the important semantic features of the sentence. Experimental results show that the proposed method performs better than the state-of-the-art methods based on convolutional neural networks and recurrent neural networks, with the  $F_1$  of 86.1% and 64.9% on the standard SemEval-2010 Task 8 and KBP37 datasets, respectively.

**Keywords:** relation extraction; neural networks; self-attention mechanism; feature-aware; multi-branch

## 0 引言

随着互联网技术的飞速发展, 如何从海量的非结构化数据中提取有用的结构化信息已成为国内外工业界和学术界的研究热点<sup>[1]</sup>。作为信息抽取

(information extraction, IE)领域的核心分支, 关系提取(relation extraction, RE)旨在从非结构化文本中提取给定实体对之间的语义关系, 目前已在知识图谱<sup>[2]</sup>、文本摘要<sup>[3]</sup>、自动问答<sup>[4]</sup>等方面得到了广泛的应用。如图 1 所示句子包含实体 1: influenza, 实体 2: virus, 目标是预测两个实体间的语义关系

收稿日期: 2020-07-18 定稿日期: 2020-08-18

基金项目: 国家自然科学基金(61673193, 62076110); 江苏省自然科学基金(BK20181341); 中国博士后科学基金(2017M621625)

Cause-Effect( $e_2, e_1$ )。

Sentence: "Avian <e1>influenza</e1> is an infectious disease of birds caused by type A strains of the influenza <e2>virus</e2>."		
Entity 1: influenza	Entity 2: virus	Relation: Cause-Effect( $e_2, e_1$ )

图1 关系提取示例

近年来,基于 Transformer<sup>[5-6]</sup>的神经网络、循环神经网络(recurrent neural network, RNN)、卷积神经网络(convolutional neural network, CNN),以及它们的变体已被广泛应用于关系提取任务<sup>[7-8]</sup>。在利用神经网络进行关系提取时,每个单词输入的向量表示是固定的,但是单词的实际含义会随着上下文的变化而改变。一些神经网络在对单词进行编码时期望能学习到其上下文的信息,然而由于没有对句子进行整体考虑<sup>[9]</sup>,难以实现学习上下文信息的目的。而 Transformer 结构中的自注意力机制(self attention mechanism, SAM)能够捕获句子内任意单词对之间的相关性特征。该方法根据词间的相关性对单词进行编码,从而使单词编码包含上下文相关性信息。RNN 利用序列单元保存时间信息,在处理时序信息时具有优势。作为典型的 RNN,双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)因其具有正反两个方向的时序结构,能够同时利用单词序列的过去和未来时刻信息,使单词的编码包含上下文时序信息。

本文首先利用 SAM 和 BiLSTM 提取单词的相关性时序特征,以对单词进行编码,使之能更精准地表示单词在句内的含义。另一方面,目前,大多数神经网络仅考虑单分支信息流,例如, Zhang 等人<sup>[10]</sup>利用 BiLSTM 对单词进行时序编码获取上下文时序信息, Nguyen 等人<sup>[11]</sup>利用不同大小的卷积核来提高 CNN 的局部感知能力。然而以上方法仍难以获取足够的语义特征来提取关系。

针对该问题,本文提出一种基于全局和局部特征感知网络(global and local feature-aware network, GLFN)的关系提取方法,在提取的相关性时序特征基础上,利用全局和局部特征感知卷积神经网络(global and local feature-aware convolutional neural network, GLFCNN)来全面表示句子的重要语义特征。具体地,针对 GLFCNN,本文首先构建多分支特征感知卷积神经网络(multibranch feature-aware convolutional neural networks, MFA-CNN),分别感知全局和局部相关性时序特征,并对这两种特征进行拼接和筛选,

最后结合 Softmax 分类器对提取到的语义信息进行关系预测。本文提出的基于全局和局部特征感知网络的关系提取方法主要包含相关性时序特征提取网络(correlated sequence feature extraction network, CSFEN),全局和局部特征感知卷积神经网络(GLFCNN),及预测输出网络(prediction output network, PON)三个部分。

(1) 相关性时序特征提取网络(CSFEN): 利用 SAM 和 BiLSTM 获取单词的相关性时序特征。具体地,前者用于学习句子中任意单词的相关性特征,后者对包含相关性特征的词进行时序编码,获取时序信息。

(2) 全局和局部特征感知卷积神经网络(GLFCNN): 首先构建多分支特征感知卷积神经网络(MFA-CNN),分别感知全局和局部相关性时序特征,并对这两种特征进行拼接和筛选,全面表示句子的重要语义特征。

(3) 预测输出网络(PON): 利用全连接层将 GLFCNN 的输出映射到分类空间,并结合 Softmax 分类器预测实体对之间的关系。

本文提出了一种基于全局和局部特征感知网络的关系提取方法,首先利用 SAM 和 BiLSTM 提取单词的相关性时序特征,并构建 GLFCNN 来获取单词的不同层次语义特征,主要贡献包含以下三点:

(1) 利用 SAM 和 BiLSTM 提取单词的相关性时序特征,更精准地表示单词在句内的语义。

(2) 构建 GLFCNN,分别感知全局和局部相关性时序特征,避免全局和局部感知的相互影响。并对感知的全局和局部特征进行拼接和筛选,以全面表示句子的重要语义特征。

(3) 在标准的 SemEval-2010 Task 8 和 KBP37 数据集上验证了所提关系提取方法的有效性,  $F_1$  分别达到 86.1% 和 64.9%。

## 1 相关工作

关系提取方面的研究主要包含早期基于人工规则的方法,以及近几年发展较快的基于神经网络的人工智能方法。前者主要利用 NLP 工具或手工设计不同的内核来选择特征<sup>[12-13]</sup>。此类方法主要存在由 NLP 工具带来的误差传播<sup>[14]</sup>,以及人工经验的局限性问题。随着神经网络的快速发展,国内外学者尝试利用其来进行关系提取。Zeng 等人<sup>[15]</sup>利用 CNN 提取词语和句子级别的特征,并结合句子中每个词与实体对之间的相对位置信息进行关系提

取,相比基于人工规则的方法更加高效。但是,受 CNN 自身结构的限制<sup>[16]</sup>,该方法在面对较远距离的实体对时,无法有效发挥其性能。RNN 擅长处理时序信息,在获取单词长距离依赖时有其优势<sup>[17]</sup>。Li 等人<sup>[18]</sup>利用 RNN 对句子进行时序编码,获取包含上下文时序信息的语义特征。Zhang 等人<sup>[10]</sup>借助 BiLSTM 强大的记忆保存能力,捕获句子中距离相隔较远的单词之间的时序依赖信息。然而,以上方法仅考虑单词的短距离依赖或时序依赖信息,没有考虑单词之间的相关性,难以描述单词在句子中的整体语义信息。本文利用 SAM 和 BiLSTM 提取单词的相关性时序特征,以对单词进行编码,使之能更精准地表示单词在句子中的上下文整体语义信息。

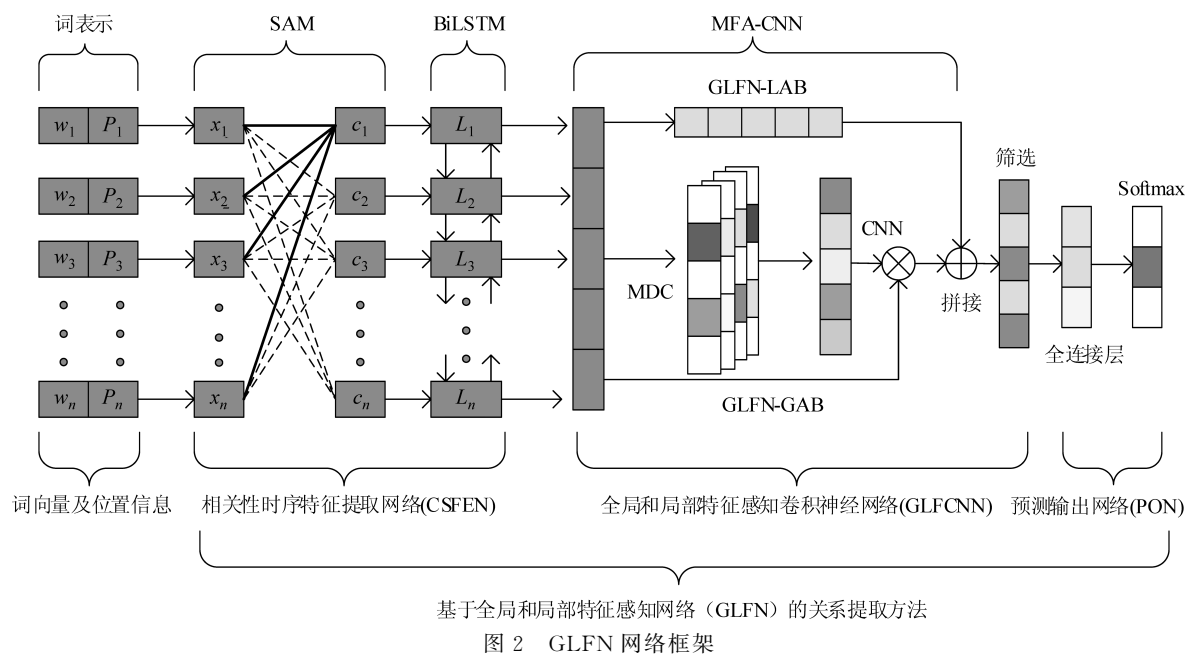
在特征感知方面,Dos 等人<sup>[7]</sup>利用 CNN 提取局部特征,并结合最大值池化操作减少冗余,感知重要特征。然而由于固定大小的卷积核会限制 CNN 的感受范围,为此,Nguyen 等人<sup>[11]</sup>利用不同大小的卷积核来提高 CNN 对局部特征的感知能力。Guo 等人<sup>[19]</sup>利用高斯注意力来提升 SAM 对局部信息的感知能力,使中心词附近的局部信息能得到更多的关

注,并进一步应用于自然语言推理任务。

然而,以上网络大多采用单分支结构,仅能感知单一的信息流特征,缺乏对句子全局和局部特征的全面关注,难以获取足够的语义特征来提取关系。针对此问题,本文提出一种基于全局和局部特征感知网络的关系提取方法,在提取的相关性时序特征基础上,进一步构建 MFA-CNN,分别感知全局和局部相关性时序特征,避免全局和局部感知的相互影响,并对感知的全局和局部特征进行拼接和筛选,以全面表示句子的重要语义。

## 2 基于全局和局部特征感知网络的关系提取方法

本节具体介绍基于全局和局部特征感知网络的关系提取方法。如图 2 所示,CSFEN 利用 SAM 和 BiLSTM 提取句子中各词之间的相关性时序特征。GLFCNN 对相关性时序特征进行全局和局部感知,获取多层次语义信息。最后通过 PON 对提取到的语义信息进行关系预测。



### 2.1 相关性时序特征提取网络(CSFEN)

CSFEN 首先采用预训练的 GloVe<sup>[20]</sup>词向量以及包含每个词与实体对的相对位置信息<sup>[15]</sup>来表示单词。因而句子的输入为  $X = \{x_1, x_2, \dots, x_n\}$ , 其中  $x_i \in \mathbb{R}^{d_x}$ 。然而,单词的实际含义往往会随着上下文的变化而发生改变,而输入的句子  $X$  是固定

的,为了使表示词的编码包含上下文相关性信息,以刻画句子的上下文语义,本文利用 SAM 学习句子内各词之间的相关性特征。SAM 主要由点乘注意力和多头注意力两个部分<sup>[5]</sup>组成。点乘注意力包含三个矩阵:询问矩阵  $Q$ 、键矩阵  $K$  和值矩阵  $V$ 。矩阵的权值由网络训练进行自动更新。点乘注意力的具体实现如式(1)所示。

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_x}}\right)\mathbf{V} \quad (1)$$

为了获取不同子空间的注意权重表示,多头注意力将注意权重映射到多个子空间中进行学习,然后将学习到的不同空间的信息进行拼接。多头注意力具体实现如式(2)、式(3)所示。

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (2)$$

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) =$$

$$\text{Concat}(\text{head}_1, \dots, \text{head}_h)\mathbf{W}^O \quad (3)$$

对于输入的句子  $X$ , 本文利用 SAM 获取包含上下文相关性信息的句子表示  $C = \text{SA}(X)$ 。SAM 用于学习句子中各词之间的相关性特征, 却忽视了句子内各词之间的时序信息。为了提取上下文正反两个方向的时序信息, 针对获取的相关性信息, 本文利用 BiLSTM 来进一步表示句子的相关性时序语义, 即输出句子的相关性时序特征  $L = \text{BiLSTM}(C)$ 。

## 2.2 全局和局部特征感知卷积神经网络(GLFCNN)

本文在获取句子相关性时序特征的基础上, 利用 GLFCNN 来全面表示句子的重要语义特征。具体地, 针对 GLFCNN, 首先构建 MFA-CNN, 包含全局特征感知分支(GLFN-GAB)和局部特征感知分支(GLFN-LAB), 分别感知全局和局部相关性时序特征。其中, GLFN-GAB 利用多通道空洞卷积(multichannel dilated convolution, MDC)获取全局特征的分布信息, 以描述相关性时序特征的重要性, 从而提取全局相关性时序特征; GLFN-LAB 则利用卷积操作获取词语的 N-gram 信息, 实现对局部相关性时序特征的感知。

GLFN-GAB 利用 MDC 将相关性时序特征映射到多个表示空间, 用以学习特征在不同空间中的分布信息。MDC 具体实现如式(4)、式(5)所示。

$$\text{Aware}_i = \text{DC}(L) \quad (4)$$

$$\text{Multichannel}(L) = \text{Concat}(\text{Aware}_1, \dots, \text{Aware}_t)\mathbf{W}^A \quad (5)$$

其中, DC 是空洞卷积操作,  $t$  是 MDC 的通道数, 决定了表示空间的数量, 本文将在实验部分讨论  $t$  的取值对网络性能的影响。在获取不同表示空间的相关性时序特征分布信息后, 本文利用高斯误差线性单元<sup>[21]</sup>(Gaussian error linear units, GELU)对分布信息进行非线性变换, 提高网络的非线性表达能力, 其计算方法如式(6)所示。

$$\text{GELU}(x) = xP(X \leq x) = x\Phi(x) \quad (6)$$

其中,  $\Phi(x)$  代表  $x$  的高斯分布的累积分布。

为了获取全局特征的分布信息, 需计算全局特征感知权重矩阵  $\mathbf{M}_{\text{aware}}$ , 其计算方法如式(7)所示。

$$\mathbf{M}_{\text{aware}}(L) = \sigma(\text{GELU}(\text{Multichannel}(L))\mathbf{W}^a) \quad (7)$$

其中,  $\sigma$  是 Sigmoid 激活函数, 用于将信息映射至 0 到 1 的范围,  $\mathbf{W}^a$  代表 GLFN-GAB 的权重。获取的  $\mathbf{M}_{\text{aware}}$  表示全局特征的重要性分布, 以感知全局相关性时序特征。

GLFN-LAB 利用  $k$  个卷积核来发掘相关性时序特征的局部 N-gram 信息  $N\text{-gram}(L)$ , 其计算方法如式(8)所示。

$$N\text{-gram}(L) = \mathbf{W}^s L + b^s \quad (8)$$

其中,  $\mathbf{W}^s$  代表卷积核的权重,  $b^s$  是卷积核的偏置, 其数值由网络训练自动更新。获取的  $N\text{-gram}(L)$  用以表示特征的局部信息, 进一步通过与全局特征  $\mathbf{M}_{\text{aware}} \odot L$  进行拼接, 融合全局与局部相关性时序特征, 其计算方法如公式(9)所示。

$$\mathbf{V} = \mathbf{M}_{\text{aware}} \odot L + N\text{-gram}(L) \quad (9)$$

其中,  $\odot$  是逐元素乘运算, 拼接的全局与局部相关性时序特征  $\mathbf{V} \in \mathbb{R}^{n \times k}$  可细化为  $\mathbf{V} = \{v_1, v_2, \dots, v_k\}$ , 其中  $v_i \in \mathbb{R}^n$  代表不同特征通道的句子表示,  $n$  和  $k$  分别表示句子的长度和特征的通道数。进一步地, 本文对得到的  $v_i$  进行最大值池化操作, 实现对特征的筛选, 每个特征通道只保留最重要的信息, 达到降低特征维度并去除冗余的目的。本文最终获取的全局与局部相关性时序特征如式(10)所示。

$$\mathbf{O} = [\max(v_1); \max(v_2); \dots; \max(v_k)] \quad (10)$$

## 2.3 预测输出网络(PON)

本文利用全连接网络将最终感知到的全局与局部相关性时序特征映射到  $|y|$  个分类标签空间, 获得预测信息  $\hat{G} \in \mathbb{R}^{|y|}$ , 如式(11)所示。

$$\hat{G} = \mathbf{W}_c \mathbf{O} + b_c \quad (11)$$

其中,  $\mathbf{W}_c$  和  $b_c$  分别代表全连接层的权重和偏置, 本文利用 Softmax 分类器对预测信息  $\hat{G}$  进行关系预测, 计算方法如式(12)所示。

$$\hat{y}_i = \frac{\exp(\hat{G}_i)}{\sum_{j=1}^{|y|} \exp(\hat{G}_j)} \quad (12)$$

其中,  $\hat{y}_i$  是实际预测输出。本文利用交叉熵作为 GLFN 的损失函数, 并采用  $L2$  正则化对网络参数加以惩罚, 以提高网络的泛化性能。网络的目标



函数如式(13)所示。

$$J(\theta) = -\frac{1}{|y|} \sum_{i=1}^{|y|} y_i \log(\hat{y}_i) + \lambda \|\theta\|_F^2 \quad (13)$$

其中,  $y_i$  是期望输出,  $\lambda$  是 L2 正则化的超参数,  $\theta$  是网络中所有的可训练参数, 由训练自动更新。神经网络的训练目标是 minimized  $J(\theta)$ , 使实际预测输出  $\hat{y}_i$  逼近于期望预测输出  $y_i$ 。

### 3 实验

#### 3.1 实验数据及评价指标

本文实验采用已完成预训练的 GloVe 词向量<sup>[20]</sup>来表示单词, 并采用 SemEval-2010 Task 8<sup>[8]</sup>和 KBP 37<sup>[22]</sup>两个标准数据集来验证所提方法的有效性。其中, SemEval-2010 Task 8 数据集包括 8 000 个训练样本和 2 717 个测试样本, 具体包含 Cause-Effect, Instrument-Agency, Product-Producer, Content-Container, Entity-Origin, Entity-Destination, Component-Whole, Member-Collection, Message-Topic, 以及 Other 共 10 种关系。由于前 9 种关系具有方向性, 本文共考虑该数据集的 19 种关系。KBP 37 数据集包括 15 917 个训练样本, 3 405 个测试样本, 具体包含 18 种具有方向性的关系, 以及一种特殊关系“No\_Relation”。因此共考虑 KBP 37 数据集的 37 种关系。此外, 本文利用宏平均  $F_1$  值来评价关系提取方法的有效性。

#### 3.2 实验结果及分析

由于 GLFN-GAB 的通道数  $t$  表示空间的数量, 选取合适的  $t$  能有效提高 GLFN-GAB 的感知能力。本文首先选取不同数量的通道数  $t$ , 分别评价仅利用全局感知分支(GLFN-GAB)和 GLFCNN (包含全局感知分支 GLFN-GAB 和局部感知分支 GLFN-LAB) 的宏平均  $F_1$  值。以上两个方法在 SemEval-2010 Task 8 和 KBP 37 数据集上的表现分别由表 1、表 2 进行展示。

表 1 SemEval-2010 Task 8 数据集上不同  $t$  值对 GLFN-GAB 和 GLFCNN 的  $F_1$  值影响

方法	通道数 $t$					
	1	2	4	8	16	32
GLFN-GAB	84.6	84.8	85.2	85.1	84.8	84.8
GLFCNN	85.6	85.7	<b>86.1</b>	85.8	85.6	85.6

表 2 KBP37 数据集上不同  $t$  值对 GLFN-GAB 和 GLFCNN 的  $F_1$  值影响

方法	通道数 $t$					
	1	2	4	8	16	32
GLFN-GAB	64.1	64.3	64.3	64.4	64.2	63.7
GLFCNN	64.4	64.5	<b>64.9</b>	64.4	64.3	63.8

由表 1 可以看出, 对于所有的通道数  $t$ , GLFCNN 的表现均优于 GLFN-GAB。由表 2 可以看出, 除开  $t=8$ , GLFCNN 和 GLFN-GAB 两者的  $F_1$  值相同外, 在其他的通道上, GLFCNN 的表现也均优于 GLFN-GAB。并且当  $t=4$  时, GLFCNN 在两个数据集上都取得最高的  $F_1$  值。

在获取最佳通道数  $t$  后, 本文还对 GLFN-GAB、GLFN-LAB、GLFCNN, 以及缺省 GLFCNN (Non GLFCNN) 的关系提取方法进行比较。图 3 和图 4 分别展示以上 4 种方法在 SemEval-2010 Task 8 和 KBP 37 数据集上的训练过程。Non GLFCNN 直接将相关性时序特征提取网络 (CSFEN) 的输出作为预测输出网络 (PON) 的输入。

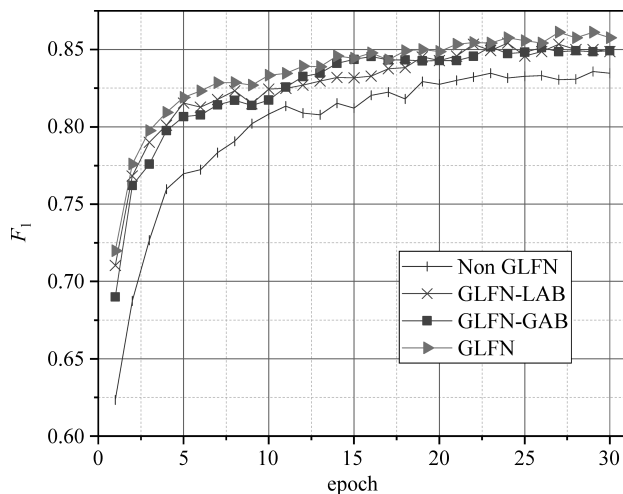


图 3 SemEval-2010 Task 8 数据集上 4 种方法的训练过程

如图 3、图 4 所示, 从总体趋势来看, Non GLFN、GLFN-LAB、GLFN-GAB 和 GLFN 这 4 种方法在开始的若干阶段随着迭代次数的增加,  $F_1$  值上升较快。而在后期若干阶段, 这 4 种方法的  $F_1$  值逐渐上升, 但也存在一定的波动。相比于 Non GLFN、GLFN-GAB 和 GLFN-LAB, GLFN 在 SemEval-2010 Task 和 KBP 37 数据集上, 分别于第 27 轮和第 38 轮获得最高  $F_1$  值。这是由于 GLFN 利用全局和局部特征感知卷积神经网络获取不同层次语义特征, 全面表示句子的重

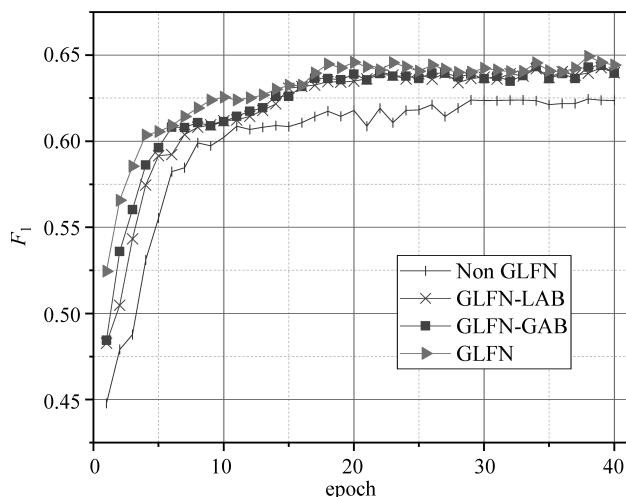


图4 KBP37数据集上4种方法的训练过程

要语义,提升了关系提取的准确性。同时,我们还利用表3和表4分别展示以上4种方法在SemEval-2010 Task 8和KBP 37数据集上的最高 $F_1$ 值。

表3 SemEval-2010 Task 8数据集上4种方法的最高 $F_1$ 值

方法	$F_1 / \%$
Non GLFN	83.6
GLFN-GAB	85.2
GLFN-LAB	85.4
GLFN	<b>86.1</b>

表4 KBP37数据集上4种方法的最高 $F_1$ 值

方法	$F_1 / \%$
Non GLFN	62.4
GLFN-GAB	64.4
GLFN-LAB	64.2
GLFN	<b>64.9</b>

由表3和表4可以看出,相比于Non GLFN、GLFN-GAB和GLFN-LAB,GLFN拥有最高的 $F_1$ 值。在SemEval-2010 Task 8和KBP37数据集上,GLFN的 $F_1$ 分别达到86.1%和64.9%。此外,表5展示了GLFN和14种主流关系提取方法(CNN、CNN+PF、Multi-CNN、CR-CNN、Attention-CNN、Bi-LSTM、BiLSTM+feature、ATT-BiLSTM、Hier-LSTM、BiLSTM-Attention、FORESTFT-DDCNN、LST-AGCN、GCN和S-Att)在SemEval-2010 Task 8数据集上的最高 $F_1$ 值。

表5 SemEval-2010 Task 8数据集上15种方法的 $F_1$ 值

方法	$F_1 / \%$
GCN <sup>[23]</sup>	79.6
Bi-LSTM <sup>[23]</sup>	80.1
CNN <sup>[23]</sup>	80.2
S-Att <sup>[23]</sup>	80.2
CNN+PF <sup>[15]</sup>	82.7
Multi-CNN <sup>[11]</sup>	82.8
CR-CNN <sup>[7]</sup>	84.1
Attention-CNN <sup>[24]</sup>	84.3
BiLSTM+feature <sup>[10]</sup>	84.3
ATT-BiLSTM <sup>[25]</sup>	84.0
Hier-LSTM <sup>[26]</sup>	84.3
BiLSTM-Attention <sup>[8]</sup>	85.2
FORESTFT-DDCNN <sup>[27]</sup>	85.5
LST-AGCN <sup>[28]</sup>	86.0
GLFN	<b>86.1</b>

最后,针对KBP 37数据集,本文还对CNN、RNN、BiLSTM-CNN、Block+CNN、Ranking CNN、Att-RCNN、Bi-SDP-Att和GLFN进行比较。表6展示了这8种方法在KBP37数据集上的最高 $F_1$ 值。

表6 KBP37数据集上8种方法的 $F_1$ 值

方法	$F_1 / \%$
CNN <sup>[29]</sup>	55.8
RNN <sup>[29]</sup>	58.8
BiLSTM-CNN <sup>[30]</sup>	60.1
Block+CNN <sup>[31]</sup>	60.9
Ranking CNN <sup>[32]</sup>	61.3
Att-RCNN <sup>[22]</sup>	61.8
Bi-SDP-Att <sup>[33]</sup>	64.4
GLFN	<b>64.9</b>

如表5所示,在SemEval-2010 Task 8数据集上,相较于14种主流关系提取方法,本文所提的GLFN具有最高的 $F_1$ 值。这是由于GLFN能分别感知全局和局部相关性时序特征,避免全局和局部感知的相互影响,并对这两种特征进行拼接和筛选,滤除贡献低的特征,以全面表示句子的重要语义,提升了关系提取的准确性。基于语法图的LST-AGCN从语法依赖关系入手来丰富句子信息,因而相对其他主流方法拥有较高的 $F_1$ 值,但该方法需借助

NLP 工具来构建语法依赖树,其性能对 NLP 工具依赖较大。如表 6 所示,在 KBP 37 数据集上,相比于 CNN、RNN、BiLSTM-CNN、Ranking CNN、Att-RCNN、Block+CNN 和 Bi-SDP-Att, GLFN 具有最高的  $F_1$  值。与  $F_1$  值次高的 Bi-SDP-Att 相比, GLFN 的  $F_1$  值提升了 0.5%,且 Bi-SDP-Att 需要构建双向短依赖路径及对应的注意力机制,其结构相对复杂。在另外 6 种方法中,Att-RCNN 拥有较高的  $F_1$  值,本文所提的 GLFN 在  $F_1$  值上较之提升了 3.1%。

#### 4 总结与展望

目前大多数神经网络仅考虑单分支信息流,难以获取足够的语义特征来进行关系提取。针对此问题,本文提出一种基于 GLFN 的关系提取方法。该方法首先利用 SAM 和 BiLSTM 获取单词的相关性时序特征。其次,构建 MFA-CNN,分别获取全局和局部相关性时序特征,避免全局和局部感知的相互影响。进一步地,对这两种特征进行拼接和筛选,以全面表示句子的重要语义特征,最后结合 Softmax 分类器实现关系提取。为了验证所提方法的有效性,在标准的 SemEval-2010 Task 8 和 KBP 37 数据集上进行广泛的实验。实验结果表明,本文所提方法的  $F_1$  值分别达到 86.1% 和 64.9%,优于主流的基于卷积神经网络和循环神经网络的关系提取方法。在今后的工作中,我们将在更多的语言词向量模型上验证所提方法的有效性。同时,我们将面向不同领域的非结构化数据,提取实体对间的关系,并构建实体关系三元组,从而建立领域结构化数据库,以进一步用于知识图谱、自动问答等任务。另外,我们还将开发便捷的用户接口,为用户提供更好的使用体验。

#### 参考文献

- [1] 袁祯祺, 宋威, 陈璟. 基于 Gate-ResNet-D 模型的远程监督关系提取方法[J]. 中文信息学报, 2019, 33(10): 57-63.
- [2] Lu R, Fei C, Wang C, et al. HAPE: A programmable big knowledge graph platform[J]. Information Sciences, 2020, 509: 87-103.
- [3] Zheng J, Zhao Z, Song Z, et al. Abstractive meeting summarization by hierarchical adaptive segmental network learning with multiple revising steps[J]. Neuro-computing, 2020, 378: 179-188.
- [4] Sun Y, Tang D, Duan N, et al. Joint learning of question answering and question generation[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 32(5): 971-982.
- [5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 5998-6008.
- [6] Wu S, He Y. Enriching pre-trained language model with entity information for relation classification[C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019: 2361-2364.
- [7] Dos C, 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, 2015: 626-634.
- [8] Lee J, Seo S, Choi Y S. Semantic relation classification via bidirectional LSTM networks with entity-aware attention using latent entity typing[J]. Symmetry, 2019, 11(6): 785-793.
- [9] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.
- [10] Zhang S, Zheng D, Hu X, et al. Bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, 2015: 73-78.
- [11] Nguyen T H, Grishman R. Relation extraction: Perspective from convolutional neural networks[C]//Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, 2015: 39-48.
- [12] Bunescu R C, Mooney R J. a shortest path dependency kernel for relation extraction[C]//Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2005: 724-731.
- [13] Mooney R J, Bunescu R C. Subsequence kernels for relation extraction[C]//Proceedings of the 18th International Conference on Neural Information Processing Systems, 2005: 171-178.
- [14] 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述[J]. 软件学报, 2019, 30(6): 1793-1818.
- [15] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network[C]//Proceed-

- ings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 2014: 2335-2344.
- [16] Shuai B, Zuo Z, Wang B, et al. Scene segmentation with dag-recurrent neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1480-1493.
- [17] Quan Z, Zeng W, Li X, et al. Recurrent neural networks with external addressable long-term and working memory for learning long-term dependencies[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31(3): 813-826.
- [18] Li J, Luong M T, Jurafsky D, et al. When are tree structures necessary for deep learning of representations? [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 2304-2314.
- [19] Guo M, Zhang Y, Liu T. Gaussian transformer: A lightweight approach for natural language inference [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 6489-6496.
- [20] Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods In Natural Language Processing (EMNLP), 2014: 1532-1543.
- [21] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.
- [22] Guo X, Zhang H, Yang H, et al. A single attention-based combination of CNN and RNN for relation classification[J]. IEEE Access, 2019, 7: 12467-12475.
- [23] Alt C, Gabryszak A, Hennig L. Probing linguistic features of sentence-level representations in relation extraction [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 1534-1545.
- [24] Shen Y, Huang X-J. Attention-based convolutional neural network for semantic relation extraction[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016: 2526-2536.
- [25] Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 207-212.
- [26] Xiao M, Liu C. Semantic relation classification via hierarchical recurrent neural network with attention [C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016: 1254-1263.
- [27] Jin L, Song L, Zhang Y, et al. Relation extraction exploiting full dependency forests [C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020: 8034-8041.
- [28] Sun K, Zhang R, Mao Y, et al. Relation extraction with convolutional network over learnable syntax-transport graph[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020: 8928-8935.
- [29] Zhang D, Wang D. Relation classification via recurrent neural network[J]. arXiv preprint arXiv: 1508.01006, 2015.
- [30] Zhang L, Xiang F. Relation classification via Bi-LSTM-CNN[C]//Proceedings of the 2018 International Conference on Data Mining and Big Data, Springer, 2018: 373-382.
- [31] Wang D, Tiwari P, Garg S, et al. Structural block driven enhanced convolutional neural representation for relation extraction[J]. Applied Soft Computing, 2020, 86: 105913.
- [32] Adilova L, Giesselbach S, Rüping S. Making efficient use of a domain expert's time in relation extraction [J]. arXiv preprint arXiv: 1807.04687, 2018.
- [33] Wang H, Qin K, Lu G, et al. Direction-sensitive relation extraction using Bi-SDP attention model[J]. Knowledge-Based Systems, 2020, 198: 105928.



宋威(1981—),通信作者,博士,教授,主要研究领域为机器学习、数据挖掘、模式识别等。  
E-mail: songwei@jiangnan.edu.cn



朱富鑫(1996—),硕士研究生,主要研究领域为自然语言处理、关系抽取等。  
E-mail: fuxin Zhu@stu.jiangnan.edu.cn