



智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

用于关系抽取的注意力图长短时记忆神经网络

张勇, 高大林, 巩敦卫, 陶一凡

引用本文:

张勇, 高大林, 巩敦卫, 等. 用于关系抽取的注意力图长短时记忆神经网络[J]. 智能系统学报, 2021, 16(3): 518–527.

ZHANG Yong, GAO Dalin, GONG Dunwei, et al. Attention graph long short-term memory neural network for relation extraction[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(3): 518–527.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202008036>

您可能感兴趣的其他文章

基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism

智能系统学报. 2020, 15(5): 956–963 <https://dx.doi.org/10.11992/tis.201903001>

基于注意力融合的图片描述生成方法

An image caption generation method based on attention fusion

智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification

智能系统学报. 2020, 15(3): 460–467 <https://dx.doi.org/10.11992/tis.201812017>

深度学习的双人交互行为识别与预测算法研究

Human interaction recognition and prediction algorithm based on deep learning

智能系统学报. 2020, 15(3): 484–490 <https://dx.doi.org/10.11992/tis.201812029>

基于双向消息链路卷积网络的显著性物体检测

Salient object detection based on bidirectional message link convolution neural network

智能系统学报. 2019, 14(6): 1152–1162 <https://dx.doi.org/10.11992/tis.201812003>

隐式特征和循环神经网络的多声部音乐生成系统

A polyphony music generation system based on latent features and a recurrent neural network

智能系统学报. 2019, 14(1): 158–164 <https://dx.doi.org/10.11992/tis.201804009>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202008036

用于关系抽取的注意力图长短时记忆神经网络

张勇, 高大林, 巩敦卫, 陶一凡

(中国矿业大学 信息与控制工程学院, 江苏 徐州 221116)

摘要: 关系抽取是信息获取中一项关键技术。句子结构树能够捕获单词之间的长距离依赖关系, 已被广泛用于关系抽取任务中。但是, 现有方法存在过度依赖句子结构树本身信息而忽略外部信息的不足。本文提出一种新型的图神经网络模型, 即注意力图长短时记忆神经网络 (attention graph long short term memory neural network, AGLSTM)。该模型采用一种软修剪策略自动学习对关系抽取有用的句子结构信息; 通过引入注意力机制, 结合句法图信息学习句子的结构特征; 并设计一种新型的图长短时记忆神经网络, 使得模型能够更好地融合句法图信息和句子的时序信息。与 10 种典型的关系抽取方法进行对比, 实验验证了该模型的优异性能。

关键词: 关系抽取; 句子结构树; 句法图; 图神经网络; 注意力图长短时记忆神经网络; 软修剪策略; 注意力机制; 长短时记忆神经网络

中图分类号: TP311 文献标志码: A 文章编号: 1673-4785(2021)03-0518-10

中文引用格式: 张勇, 高大林, 巩敦卫, 等. 用于关系抽取的注意力图长短时记忆神经网络 [J]. 智能系统学报, 2021, 16(3): 518-527.

英文引用格式: ZHANG Yong, GAO Dalin, GONG Dunwei, et al. Attention graph long short-term memory neural network for relation extraction[J]. CAAI transactions on intelligent systems, 2021, 16(3): 518-527.

Attention graph long short-term memory neural network for relation extraction

ZHANG Yong, GAO Dalin, GONG Dunwei, TAO Yifan

(School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: Relation extraction is a key technology in information acquisition. The sentence structure tree that can capture long-distance dependencies between words has been widely used in relational extraction tasks. However, existing methods still have the disadvantage of relying too much on the information of sentence structure tree and ignoring external information. This paper proposes a new graph neural network structure, namely the attention graph long short term memory neural network (AGLSTM). The model adopts a soft pruning strategy to automatically learn sentence structure information useful for relation extraction tasks; then the attention mechanism is introduced and combined with the syntactic graph information to learn the structural features of the sentence; And designed a new type of graph long short term memory neural network to better fuse syntactic graph information and sentence timing information. Compared with 10 typical relational extraction methods, experiments verify the excellent performance of the proposed method.

Keywords: relation extraction; sentence structure tree; syntactic diagram; graph neural network; AGLSTM; soft pruning strategy; attention mechanism; LSTM

关系抽取是信息抽取的一个重要组成部分, 已被成功用于医学知识发现^[1]、知识图谱构建^[2]和问答推理^[3]等诸多自然语言处理问题。关系抽取是指从非结构化的文本中抽取实体对之间的三元组信息。

现有关系抽取模型或方法可大体分为 2 类:

收稿日期: 2020-08-30.

基金项目: 国家重点研发计划项目 (2020YFB1708200); 科技部科技创新 2030 重大项目 (2020AAA0107300).

通信作者: 高大林. E-mail: 1367963012@qq.com.

序列模型和依赖关系模型。序列模型主要针对单词序列进行建模^[4-5], 而依赖关系模型则是以句法依存关系树为基础, 构建依赖关系模型^[6-7]。已有研究表明, 相对序列模型而言, 依赖关系模型可以利用句法依存树提取出句子中隐藏的非局部语法关系^[8-9], 进而建立句内长距离的依赖关系。然而, 在生成句法依存树时, 它常依赖句法结构规则对句子结构进行修剪。这种基于规则的修剪策略可能会删除句子结构所包含的重要隐式信息。

为此,Guo等^[10]提出了一种基于软修剪的关系提取模型,用以实现模型自主学习句法结构,而无需对句法结构进行硬性修剪。虽然所提模型可以很好地学习句法结构,但是,由于图卷积模型本身的无向性,其对上下文中时序特征的学习能力较弱。

目前,已有学者以句法依存树为基础,提出了多种图神经网络模型,但是所提模型大多是基于图卷积神经网络^[11]。图卷积神经网络可以很好地学习图结构数据的信息,却难以有效处理时序数据。这就意味着,面向具备时序特性的文本数据,只依赖图卷积神经网络并不能很好地抽取文本中时序信息。相对图卷积神经网络,长短时记忆神经网络(long short term memory neural network, LSTM)可以有效提取上下文的时序信息。目前,部分学者开始尝试将其用于跨句关系的抽取问题,并提出了多种图长短时记忆神经网络^[12-13]。但是,这些方法都是基于修剪后句子结构树的信息来构建图神经网络,依然存在误删有用信息的不足。另外,传统方法对句子结构信息的使用完全依赖于句子结构树。如果句子结构树中蕴含错误信息,很容易累加错误信息。

鉴于此,本文提出一种注意力图长短时记忆神经网络(attention graph long short term memory neural network, AGLSTM)。采用注意力机制与句子结构树相结合的方式建立句子的全局结构信息,生成全局注意力权重矩阵,并将其用于构造AGLSTM模型。所提方法不仅可以充分利用LSTM学习句子的时序结构特征,而且可以根据句法图信息自行学习句子的结构特征。相对于完全依赖句子结构树的传统方法,本文方法是在原始句子结构树的基础上,利用注意力机制使得模型自行学习句子结构中的信息,不仅可以得到更加有效的关系抽取模型,而且鲁棒性更好。

1 相关工作

关系抽取是自然语言处理中的一项重要分支任务,在从非结构化文本中抽取结构性信息时起着重要作用。现有主流关系抽取技术可分为2种:远程监督关系抽取^[14-15]和有监督关系抽取^[5,16]。

有监督关系抽取法把关系抽取看作一种多分类任务进行处理。早期的研究成果主要是使用核函数等机器学习方法设计分类器^[17],并进行关系抽取任务。随着深度学习的兴起,深度神经网络已在关系抽取中得到了广泛应用。Zeng等^[18]采用深度卷积神经网络(convolutional neural network,

CNN)提取词汇级和句子级特征,进而完成关系抽取;Zhang等^[16]利用长短时记忆神经网络学习实体对之间的长期依赖性;Zhang等^[19]在词嵌入的基础上考虑单词的位置和词性、句子结构依赖等信息,设计了一种新型的长短时记忆神经网络进行关系抽取。马语丹等^[20]通过引入注意力机制提取实体的网络环境信息,生成全局的上下文特征。鉴于LSTM对文本信息建模的优越性,以及句子结构信息对文本关系抽取任务的重要性,学者们还提出了结合句子结构信息建模的SDP-LSTM^[6]和Tree-LSTM^[7]等改进模型。

远程监督关系提取方法借助知识库,利用知识库中包含的知识来对非结构化文本进行标注。这种方法可以自动构建大量训练数据,不仅解决了现有方法数据量不足的缺点,而且增强了模型的跨领域适应能力。Mintz等^[21]最早提出了远程监督的基本假设,定义了远程监督任务。然而,远程监督在解决数据量不足的同时,容易带来较大的噪声。为此,越来越多的学者开始关注对远程监督方法的改进研究。Zeng等^[22]融合分段卷积神经网络和多实例学习(multiple instance learning, MIL),提出了一种混合关系提取模型。该模型通过PCNN提取句子特征,利用MIL消除错误样本噪声,进而实现远程监督关系抽取。

2 所提关系抽取模型

2.1 模型框架

为同时学习句法图中句子的结构特征和句子之间的时序结构特征,提出一种新的图神经网络结构,即注意力图长短时记忆神经网络(AGLSTM)。图1给出了融合AGLSTM模型的关系抽取框架。

2.2 Embedding层

关系抽取方法需要考虑语料中的实体语义特征和实体位置特征。在自然语言处理领域中,词向量表示常用于表示词的语义信息。本文通过词嵌入方法对整个语料库进行训练,得到语料库的词向量矩阵,其中每个词表示为一个 d_{word} 维的向量 \mathbf{w}_{word} 。另外,为了更好地表示实体的语义特征,利用stanfordparser工具对语料库进行命名实体识别(NER)分析,将生成的NER标签映射成一个 d_{ner} 维的向量 \mathbf{w}_{ner} 。将 \mathbf{w}_{word} 和 \mathbf{w}_{ner} 连接即为词的实体表示向量。

本文采用Zeng等^[18]提出的实体位置嵌入方法引入实体位置信息。将句中每一个单词与实体对 (e_1, e_2) 的相对位置 (p_1, p_2) 映射成 d_{pos} 维的距离

向量 (w_{pos}^1, w_{pos}^2) , 其中 w_{pos}^1 表示单词对实体 e_1 的距离信息, w_{pos}^2 表示单词对实体 e_2 的距离信息。最终实体的位置特征表示为 $w_{pos} = [w_{pos}^1, w_{pos}^2]$ 。

将实体语义特征嵌入与实体位置特征嵌入融合, 作为模型嵌入层的输出, 即 $w_u = [w_{word}; w_{ner}; w_{pos}]$, 向量长度为 $d_w = d_{word} + d_{ner} + 2d_{pos}$ 。

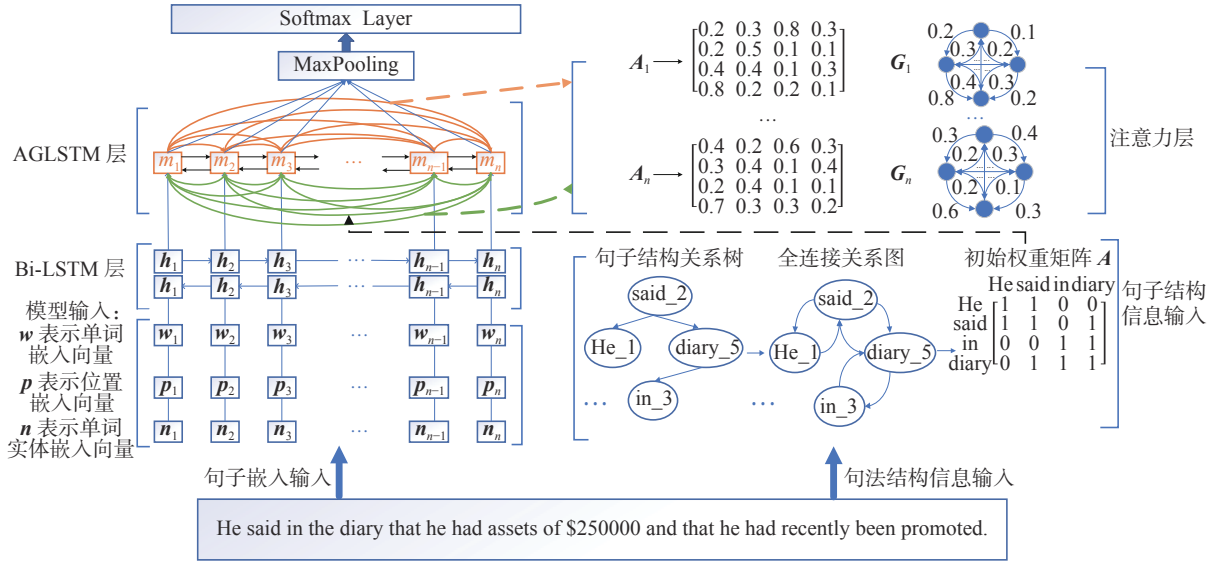


图1 AGLSTM模型的关系抽取框架

Fig. 1 Relation extraction framework based on the AGLSTM model

2.3 Bi-LSTM层

嵌入层输出的特征矩阵 w_u 包含每一个单词本身的信息, 句子中每个词所包含的信息均独立。因为句子结构信息包含单词与单词间的相互关系, 所以, 为更好地将数据本身与句子结构信息相结合, 需要先通过一层双向 LSTM 层来获取词与词之间的语义连接, 方法如下:

$$\begin{aligned}\vec{h}_t &= \overrightarrow{\text{LSTM}}(w_u, \vec{h}_{t-1}) \\ \overleftarrow{h}_t &= \overleftarrow{\text{LSTM}}(w_u, \overleftarrow{h}_{t-1})\end{aligned}$$

最终, 将 Bi-LSTM 的正向和反向隐藏层输出相连接, 可以得到输出特征 $h_t = \vec{h}_t \oplus \overleftarrow{h}_t$ 。

2.4 句法结构矩阵的生成

对文本数据进行语义信息抽取时, 不仅要考虑句子所包含单词的特征信息, 而且需要挖掘句子本身所蕴含的结构信息。已有大量研究成果将句法结构用于关系抽取, 取得了很好的效果。

为建立句法结构矩阵, 首先利用句法依存分析提取出每个句子序列 $S = \{S_1, S_2, \dots, S_n\}$ 的句法结构。例如, 对句子 “He said in the diary that he had assets of \$250 000 and that he had recently been promoted.” 进行句法分析, 图 2 给出了它的句法树。接着, 将提取出的句法树信息转换成全连接图的矩阵形式。图 1 右侧句子结构信息输入部分展示了构建初始邻接权重矩阵 A 的过程, 其中, 全连接关系图展示了扩展后的依存树。进一步, 将依存关系图写成矩阵形式, 得到初始的句法结构邻接矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

如图 1 右侧句子结构信息输入部分中初始权重矩阵所示。将得到的邻接矩阵 A 与经过 Bi-LSTM 得到的文本特征相结合, 输入到所提模型的 AGLSTM 层。

2.5 注意力图长短时记忆神经网络 (AGLSTM)

现有方法大都依据句法结构信息将句子处理成邻接子树, 如图 2 所示。因为在子树中只有相连叶子结点的单词可以构成联系, 这些方法有可能丢失远程单词之间的关联信息。鉴于此, 本文采用软修剪策略, 将句子的邻接子树转换为全连接的边加权图; 并通过注意力的形式, 使模型可以根据初始邻接矩阵信息自行训练注意力权重, 进而更全面地学习句子的结构信息。

AGLSTM 是由图长短时记忆神经网络 (graph long short term memory neural network, GLSTM) 和注意力机制结合而成, 如图 1 左侧所示。GLSTM 中每一个单词节点 w_j 的表达包括: 单词节点的输入特征向量表示 x_j , 句法图内所有与该单词相连的边的表达 h_j^{in} 和 h_j^{out} 。每一个边的表达是通过相连神经元节点的特征与边权重乘积得到, 网络结构如图 3 所示。GLSTM 中边权重只有 1 或 0 两种状态, 当为 0 时代表该边不存在, 而 AGLSTM 是采用全连接的结构, 边权重为 0~1 中的归一化值, 因此学习到的信息更加丰富。

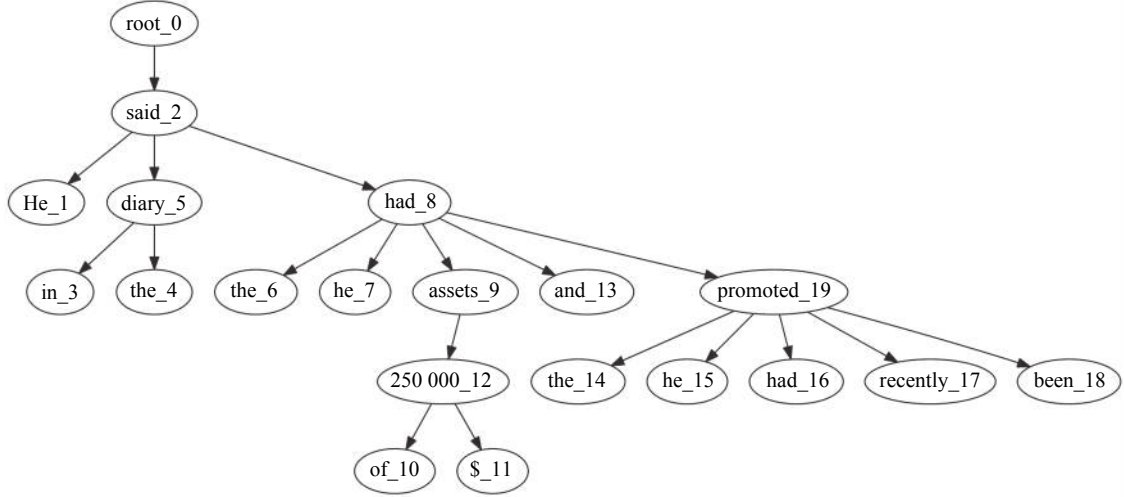


图2 句法分析树示例

Fig. 2 Schematic diagram of the syntactic analysis

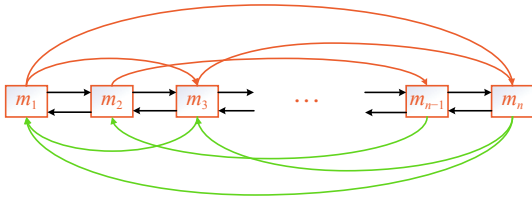


图3 GLSTM 网络结构

Fig. 3 Structure of the GLSTM network

在AGLSTM中每2个单词节点之间的边权重为注意力层所得多头注意力矩阵中相应元素的权重值。 h_j^{in} 和 h_j^{out} 的计算公式分别为

$$h_j^{\text{in}} = \sum_{(i,j) \in \text{all}} \alpha_{ij} h_i \quad (1)$$

$$h_j^{\text{out}} = \sum_{(j,k) \in \text{all}} \alpha_{jk} h_k \quad (2)$$

式中: α_{ij} 为从节点*i*到节点*j*的句子结构权重; h_j^{in} 表示节点*j*的所有边的输入; h_j^{out} 表示节点*j*所有边的输出; h_i 和 h_k 为节点*i*和*k*的隐藏状态特征。GLSTM网络定义为

$$i_j = \sigma(W_i x_j + U_{i1} h_j^{\text{in}} + U_{i2} h_j^{\text{out}} + b_i) \quad (3)$$

$$o_j = \sigma(W_o x_j + U_{o1} h_j^{\text{in}} + U_{o2} h_j^{\text{out}} + b_o) \quad (4)$$

$$f_j = \sigma(W_f x_j + U_{f1} h_j^{\text{in}} + U_{f2} h_j^{\text{out}} + b_f) \quad (5)$$

$$u_j = \sigma(W_u x_j + U_{u1} h_j^{\text{in}} + U_{u2} h_j^{\text{out}} + b_u) \quad (6)$$

$$c_j = f_j \odot c_{j-1} + i_j \odot u_j \quad (7)$$

$$h_j = o_j \odot \tanh(c_j) \quad (8)$$

式中: i_j 、 o_j 和 f_j 分别是输入门、输出门和遗忘门; W_x 、 U_x 和 $b_x(x \in (i, o, f, u))$ 为模型参数; h_j 为节点*j*的隐藏状态特征。

结合注意力机制和GLSTM,构造注意力图长短时记忆神经网络(AGLSTM)。利用注意力机制对句子结构特征进行注意力运算,得到注意力权

重矩阵 A^k 。 A^k 中每一项代表节点*a*到*b*的边权重。采用多头注意力机制(multi-head-attention),使模型从不同表征子空间中获得多层面的信息,最终可得到*k*个注意力权重矩阵:

$$A^k = \begin{bmatrix} \alpha_{11}^k & \alpha_{12}^k & \cdots & \alpha_{1n}^k \\ \vdots & \vdots & & \vdots \\ \alpha_{n1}^k & \alpha_{n2}^k & \cdots & \alpha_{nn}^k \end{bmatrix}, \quad k = 1, 2, \dots, n$$

式中:每个矩阵 A^k 代表句子结构的一个权重图G。图1中右侧注意力层给出了注意力层输出的*k*个矩阵 $A^k(k = 1, 2, \dots, n)$,每一个注意力权重矩阵 A^k 分别与权重图G一一对应。最后,将*k*个注意力权重矩阵进行线性变换,整合成最终的权重矩阵A,即为AGLSTM层的重要参数。

Self-attention机制可以交互计算序列中任意位置,从而捕获整个序列内的依赖关系。具体地,Self-attention计算公式为

$$u_i = \tanh(W_u H_{\text{LSTM}} + b_u)$$

$$A^k = \text{softmax}(W_u u_i)$$

式中: H_{LSTM} 是Bi-LSTM层提取的文本特征向量; u_i 是以tanh为激活函数的感知机层得到的隐藏特征向量。通过softmax函数获得标准化的权重矩阵 A^k 。经过*k*次self-attention计算,可以得到多头注意力机制输出的*k*个注意力权重矩阵 A^k 。

将得到的注意力权重矩阵 A^k 代入式(1)、(2),计算单词节点 w_j 的隐藏状态;再将得到的 h_j^{in} 和 h_j^{out} 用于式(3)~(8)中,通过注意力机制与GLSTM的网络结构相结合,进而构造AGLSTM层。

2.6 线性连接层

根据多头注意力层,可以得到*k*个不同的注意力权重矩阵A。AGLSTM层输出便得到*k*个不同的结果。本文利用线性连接层将这*k*个不同的

结果整合成最终的输出特征。线性连接层定义的数学形式为

$$\mathbf{h}_{\text{comb}} = \mathbf{W}_{\text{comb}} \mathbf{h}_{\text{out}} + \mathbf{b}_{\text{comb}}$$

式中: \mathbf{W}_{comb} 是线性连接层的权重矩阵; \mathbf{b}_{comb} 是线性连接层的偏置矩阵; \mathbf{h}_{comb} 为最终输出的特征。

2.7 输出层

针对最终输出的特征 \mathbf{h}_{comb} , 先通过最大池化层, 保留句子中最重要的特征, 再通过 softmax 函数计算每一个类别的条件概率, 进而输出模型预测特征的标签类别。每一个特征对应标签类别的预测概率为

$$p(h_i) = \frac{h_{\text{comb}}^i}{\sum_{n=1}^N h_{\text{comb}}^n}, \quad i = 1, 2, \dots, N$$

3 实验分析

3.1 数据集

本节在 2 个常用的关系抽取数据集中验证所提模型的有效性。它们分别是 TACRED 数据集^[19]和 SemEval 2010 task 8 数据集^[23]。

TACRED 数据集: 该数据集共 106 264 条实例, 分为训练集 68 124 条, 验证集 22 631 条, 测试集 15 509 条。数据集引入了 41 种关系类型和一种特殊的“no relation”类型, 来描述实例中相关实体对之间的关系。

SemEval 2010 task 8 数据集: 该数据集主要用于对名词性实体之间的语义关系进行分类。它采用 9 种关系类型和一种特殊的关系类型“Other”来描述实体对之间的关系。数据集一共包含 10 717 条实例, 其中, 8 000 条实例为训练数据, 其余为测试数据。

3.2 模型参数与性能指标

采用 Adam 算法优化本文所提模型, 网络隐藏层节点数设置为 300。为防止过拟合, 在输入层和 AGLSTM 层同时引入损失率 (dropout), 并且在目标函数中引入 L2 正则化项, 正则化因子取值为 0.001。对于 TACRED 数据集, 模型参数设置学习率为 0.7, 在输入层和 AGLSTM 层的损失率都设为 0.5。对于 SemEval 2010 task8 数据集, 模型设置学习率为 0.5, 在输入层和 AGLSTM 层的损失率分别设为 0.5 和 0.3。采用预训练好的 Glove 向量^[24]对数据集进行词向量训练, 生成词嵌入向量; 结合实体位置嵌入和 NER 标签嵌入, 生成最终嵌入层向量。

在注意力层, 多头注意力机制的参数 k 会对模型性能产生较大的影响。以 SemEval 数据集为

例, 依次设置 $k=[2,3,4,5,6]$, 下面讨论 k 值对所提模型性能的影响。表 1 展示了不同 k 值下所提模型的 F_1 指标值。可以看出, 当 $k=4$ 时, 模型得到了最佳的 F_1 指标值。鉴于此, 后续实验中设置 $k=4$ 。

表 1 不同 k 值下所提模型的实验结果

k 头数	F_1	%
$k=2$	84.6	
$k=3$	85.1	
$k=4$	85.3	
$k=5$	84.7	
$k=6$	84.6	

3.3 评价指标

本文采用关系抽取任务中常用的 3 类评价指标: 精确率 P 、召回率 R 和 F_1 值。二分类中常用的计算公式为

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

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

$$F_1 = \frac{2PR}{P+R}$$

式中: TP(true positives) 表示将正样本预测为正的样本数; FP(false positives) 表示将负样本预测为正的样本数; FN(false negatives) 表示将正样本预测为负的样本数。

3.4 对比算法

将 AGLSTM 模型与 10 种典型关系抽取方法进行对比, 它们分别为

- 1) 逻辑回归模型 (logistics regression, LR)。
- 2) 支持向量机模型 (support vector machine, SVM)。

3) 双向 LSTM 模型+注意力机制 (BiLSTM+Att)^[4]: 该模型将双向 LSTM 模型与注意力机制结合进行关系抽取。

4) 位置感知注意力 LSTM 模型 (PA-LSTM)^[19]: 该模型通过对引入实体位置的注意力机制建立模型。

5) 最短依存路径 LSTM 模型 (SDP-LSTM)^[6]: 该模型使用 2 个实体之间的最短依存路径, 并结合 LSTM 模型进行建模。

6) 树结构 LSTM 模型 (tree-LSTM)^[7]: 不同于以往的顺序 LSTM 模型, 该模型结合句子依存关系树将 LSTM 推广到树形结构。

7) 分段注意力 LSTM 模型 (SA-LSTM)^[25]: 考虑到部分实体关系信息包含在句子中的某些片段

的情况,该模型设置了分段注意力层来进行这种信息的提取,给出了分段注意力 LSTM 模型。

8) 实体感知注意力 LSTM 模型 (EA-LSTM)^[26]: 结合了实体感知注意力和潜在实体类型,该模型给出了一种新的端到端的神经网络模型。

9) 图卷积神经网络模型 (GCN)^[8]: 该模型提出了一种以路径为中心的剪枝策略。结合图卷积网络,在最大限度保留有关内容的同时,GCN 能从树中删除无关信息,进一步提高关系抽取的性能。

10) 注意力导向图卷积神经网络模型 (AG-GCN)^[10]: 该模型给出了一种基于注意力机制的软修剪策略,并采用图卷积进行建模。

在上述 10 种对比方法中,LR 与 SVM 是基于机器学习的模型;BiLSTM+Att、PA-LSTM、SDP-LSTM、SA-LSTM 和 EA-LSTM 隶属于基于顺序结构的神经网络模型;Tree-LSTM、GCN 与 AG-GCN 为基于依赖结构的神经网络模型。本文通过与这 3 大类模型进行对比,验证所提模型的有效性。

3.5 结果分析

3.5.1 数据集 TACRED

表 2 展示了 AGLSTM 和对比算法处理该数据时得到的 3 个性能指标值。可以看出:与基于机器学习的模型 LR 相比,在 3 个指标 (P 、 R 和 F_1) 上,所提 AGLSTM 模型都获得了最优的指标值。特别地,AGLSTM 所得 R 和 F_1 值明显高于 LR 所得值。

表 2 TACRED 数据集实验结果
Table 2 Experimental results on TACRED %

模型	P	R	F_1
LR	73.5	49.9	59.4
PA-LSTM	65.7	64.5	65.1
SDP-LSTM	66.3	52.7	58.7
SA-LSTM	68.1	65.7	66.9
Tree-LSTM	66.0	59.2	62.4
GCN	69.8	59.0	64.0
AGGCN	69.9	60.9	65.1
AGLSTM(ours)	74.0	62.2	67.5

与 3 种基于顺序结构的神经网络模型 (PA-LSTM、SDP-LSTM 和 SA-LSTM) 相比,AGLSTM 获得了明显优于 3 种算法的 P 指标值。以 PA-LSTM 为例,AGLSTM 的 P 指标值提升了 8.3;在召回率 R 指标上,SA-LSTM 和 PA-LSTM 的结果要好于本文所提模型 AGLSTM。然而,从综合指标 F_1 值来看,AGLSTM 的结果要明显优于 3 种算法。相对 PA-LSTM、SDP-LSTM 和 SA-LSTM,AGLSTM 的 F_1 指标值分别提升了 2.4、8.8 和

0.6。这主要因为,相比基于顺序结构的神经网络模型,AGLSTM 可以同时捕捉文本中顺序结构和依赖结构信息。

与基于依赖结构的神经网络模型 Tree-LSTM 和 GCN 相比,AGLSTM 模型的 F_1 值分别提高了 5.1 和 3.5。这表明,在捕获复杂语法特性模式、建立句子长期依赖关系模型方面,AGLSTM 的能力要优于两种对比算法。与 AGGCN 模型相比,AGLSTM 模型的 F_1 值提高了 2.4。这说明,AGLSTM 在满足捕获句子结构信息的同时,也能很好地捕获上下文的时序信息。另外,在精确率 P 上,本文所提 AGLSTM 模型的结果也要明显高于其他模型。综上所述,相比于其他对比算法,AGLSTM 模型是一种极具竞争力的关系抽取方法。

3.5.2 数据集 SemEval 2010 task 8

将 AGLSTM 模型与多种典型对比算法用于该数据集,表 3 给出了它们所得的 F_1 指标值。一方面, F_1 指标是对精确率 P 和召回率 R 的统一融合,可以更为全面地反映算法的性能;另一方面,相关对比文献仅仅给出了该数据集的 F_1 值。鉴于此,本文也只使用 F_1 值作为评价指标。

表 3 SemEval 数据集实验结果
Table 3 Experimental results on SemEval %

模型	F_1
SVM	82.2
CNN+Att	84.1
BILSTM+Att	84.0
SDP-LSTM	83.7
PA-LSTM	82.7
SPTree	84.4
C-GCN	84.8
AGLSTM(ours)	85.3

相比 TACRED 数据集,SemEval 数据集的样本规模较小,共有 10 717 条数据。通过表 3 可知,所提 AGLSTM 模型在小规模数据集 SemEval 上也能取得较好的效果。与基于机器学习的模型 SVM 相比,7 种基于神经网络模型的抽取算法都获得了较好的 F_1 值。这说明,复杂语义信息的提取对语言模型的建立有很大帮助,也对关系抽取的结果产生了很大影响,而神经网络模型更适合进行语义信息的建模。与基于顺序结构的神经网络模型 (CNN+Att、Bilstm+Att、SDP-LSTM 和 PA-LSTM) 相比,AGLSTM 所得 F_1 值分别提高了 1.2、1.3、1.6 和 2.6;与基于依赖结构的神经网络模型 SPTree 和 GCN 相比,AGLSTM 模型的 F_1 值分别提高了 0.9 和 0.5。这主要因为,相比这些对比

算法, AGLSTM可以同时捕捉文本中顺序结构和依赖结构等2方面信息。

3.6 消融分析

表4和表5分别展示了AGLSTM模型在2个数据集上消融分析的结果。针对TACRED数据集,由表3可知:1)删除LSTM层时,AGLSTM的 F_1 值下降了1.8。这说明,利用LSTM对特征进行初次提取对后续GLSTM层的处理有很大的帮助。2)当删除注意力层时,AGLSTM模型的 F_1 结果下降了1.7。这意味着,注意力层在模型机构中起到很重要的作用。3)当移除句法依存信息时, F_1 结果下降了0.5。本文所提模型是以句子结构为基础实现的注意力机制,所以当除去句子结构信息时,模型使用的注意力机制退化为依据句子顺序序列实现的传统注意力机制。可见,句子结构信息在模型中也起到了较为重要的作用。

表4 TACRED数据集上的消融研究结果

Table 4 Ablation analyses on TACRED %

模型	F_1
删除句法依存信息	67.0
删除LSTM层	65.7
删除注意力层	65.8
AGLSTM	67.5

表5 SemEval数据集上的消融研究结果

Table 5 Ablation analyses on SemEval %

模型	F_1
删除句法依存信息	85.0
删除LSTM层	84.5
删除注意力层	84.2
AGLSTM	85.3

类似地,在对SemEval数据集进行分析时发现,当分别去掉句法依存信息、LSTM层、注意力层时,AGLSTM的 F_1 指标值也均有所下降。这些都说明,上述3个结构在模型训练中起了重要的作用。

采用配置为2080Ti的单个GPU进行模型训练,图4、5展示了所提模型在处理数据集TACRED和Semeval时消融分析的训练时长。可以看出,针对2个数据集,AGLSTM模型的运行花费分别为653 min和76 min;当分别除去LSTM层、句子结构信息和注意力机制时,模型的训练时间均有所降低;特别地,当除去注意力机制时,AGLSTM运行时间的减少幅度最为明显。这说明,注意力机制确实对模型复杂度有较大影响,同时它也显著提高了模型的性能。此外,多

头注意力机制中的 K 值也会影响模型的复杂度。当 K 从2变化到5时,模型训练的时间呈现递增趋势。

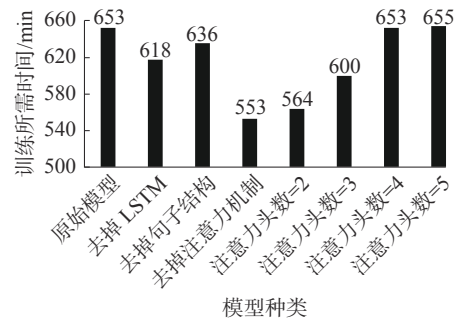


图4 TACRED数据集下模型训练时间分析

Fig. 4 Model training time analysis under the TACRED dataset

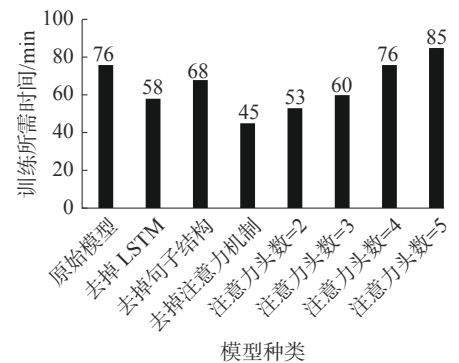


图5 Semeval数据集下模型训练时间分析

Fig. 5 Model training time analysis under the Semeval dataset

3.7 模型的泛化性分析

从表2和表3的实验结果可知,相对其他对比模型,在采用相同训练集和测试集的情况下,所提模型在数据集TACRED和SemEval都取得了较高的 F_1 值。这在一定程度上说明所提模型的泛化性好于其他对比模型。针对不同句子长度进行模型泛化性分析,将数据集TACRED划分成不同句子长度的数据集,图6展示了不同句子长度下所提模型的 F_1 值。可以看出:1)当句子长度位于20~60时,所提模型的 F_1 值变化不大,而其他对比模型的 F_1 值变化幅度都要大于本文所提模型。这说明本文所提模型更加稳定,具有更好的泛化性和鲁棒性;2)与SDP-LSTM、PA-LSTM和LSTM等基于顺序神经网络的模型相比,在处理较短句子时,基于图结构神经网络的模型(AGLSTM和GCN)的性能表现略差;但是,对于复杂长句子,AGLSTM和GCN的性能远高于3种基于顺序神经网络的模型。这表明,相对基于顺序神经网络的模型,基于图神经网络的模型可以更好地处理复杂句子。

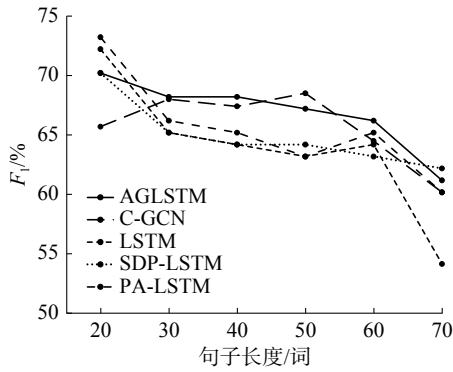


图6 句子长度分析结果

Fig. 6 Results of the sentence length analysis

3.8 可视化分析

本节通过对模型中使用的注意力机制进行可视化分析,说明所提模型更具可解释性。注意力机制的可视化显示了句中每个单词对于其他词的重要程度,可以通过注意力机制来获得更丰富的词汇表征。图7展示了句子“In an August 2004 Fortune article, Countrywide CEO Angelo Mozilo was mocking WaMu for its stumbles in the mortgage business”

business.”的可视化结果。

图7中4个子图分别对应4个注意力机制的头。本文通过全连接构建完整的句子结构,因此每个单词对句子中所有单词均有一个权重,图7中颜色的深浅代表关系权重的高低,即单词之间相关性的强弱。可以看出,图7中对角线的颜色普遍较深,说明每个单词对应自己的权重最大;“Countrywide”和“Angelo Mozilo”是一对实体对,“Countrywide”是一个机构名,它的权重主要集中在“CEO”和“article”等相近的词上;“Angelo Mozilo”是一个人名,它的注意权重主要集中在“CEO”、“was”和“mocking”等表现人物特征的词上。实际上,这些词对“Countrywide”和“Angelo Mozilo”之间关系“org:top_members/ employees”有很重要的预测作用。因此这些词学习到的注意力权重较大,颜色较深;其他单词对于句子结构信息的建模也起着一定的作用,不能完全忽略,其颜色较浅。这些都表明,模型使用的注意力机制可以更好地建模句子的全局结构信息。

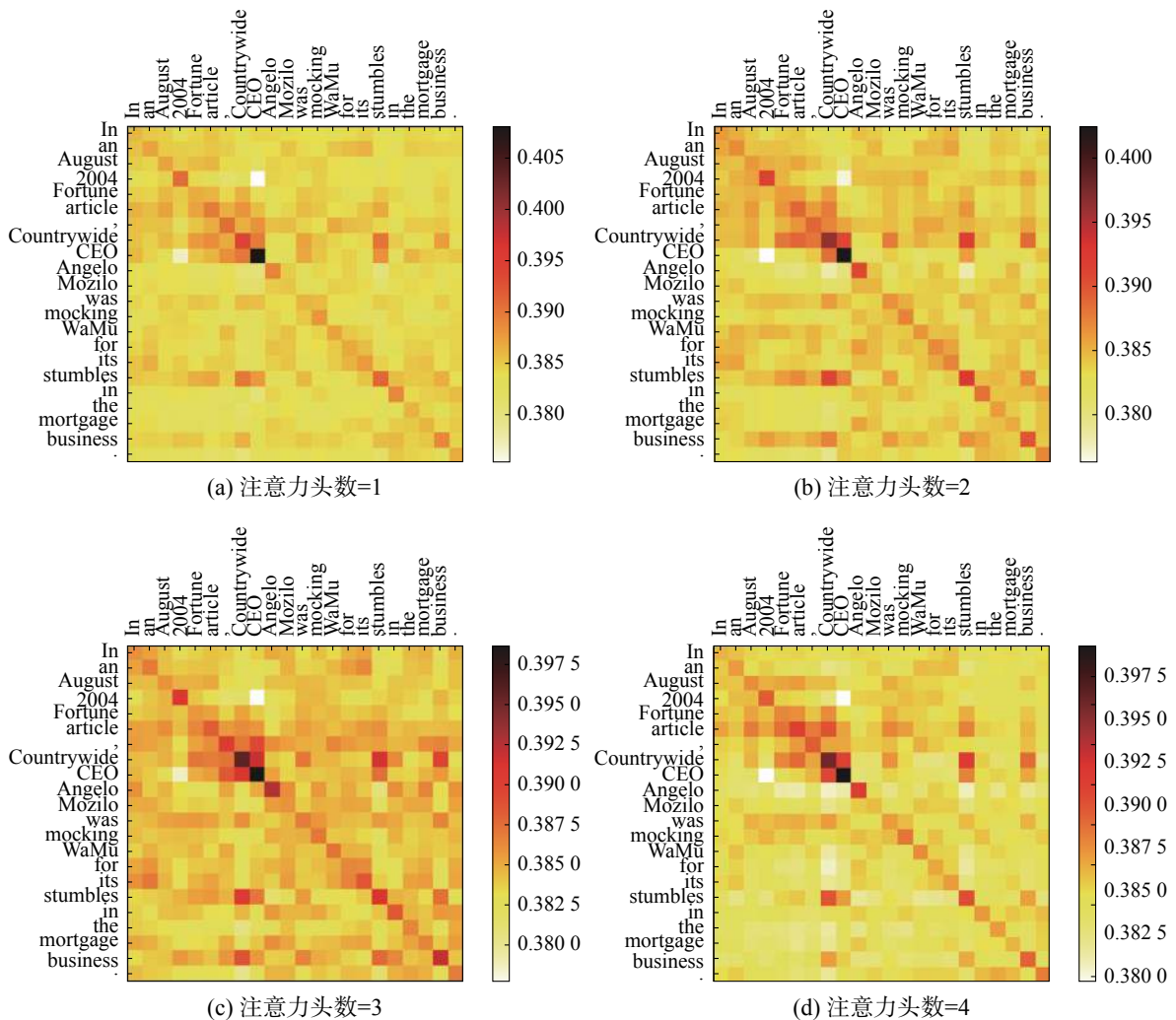


图7 注意力机制可视化分析

Fig. 7 Visual analysis of the attention mechanisms

4 结束语

本文提出了一种新型的注意力图长短时记忆神经网络模型 (AGLSTM)。该模型将注意力机制与句子结构树结合, 实现了模型自主学习句子结构信息的功能。所提模型不仅拥有很好的捕捉复杂语义关系和依赖结构的能力, 并且弥补了图卷积网络对时序信息捕捉能力差的不足。与 10 种关系提取模型或方法进行对比, 实验结果表明, 所提模型在关系抽取上具有较佳的性能, 其准确率要远高于其他对比模型。在未来工作中, 将深入研究句内结构以及句间结构的信息提取, 将模型进行完善并应用到句间关系抽取任务。

参考文献:

- [1] 杨志豪, 洪莉, 林鸿飞, 等. 基于支持向量机的生物医学文献蛋白质关系抽取 [J]. 智能系统学报, 2008(4): 361–369.
Yang Zhihao, Hong Li, Lin Hongfei, et al. Extraction of information on protein-protein interaction from biomedical literatures using an SVM[J]. CAAI transactions on intelligent systems, 2008(4): 361–369.
- [2] 李智超. 图文知识图谱中的关系抽取算法研究 [D]. 北京: 北京邮电大学, 2018.
LI Zhichao. A relation extraction algorithm in multi-modal knowledge graph[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.
- [3] 张涛, 贾真, 李天瑞, 等. 基于知识库的开放领域问答系统 [J]. 智能系统学报, 2018, 13(4): 557–563.
ZHANG Tao, JIA Zhen, LI Tianrui, et al. Open-domain question-answering system based on large-scale knowledge base[J]. CAAI transactions on intelligent systems, 2018, 13(4): 557–563.
- [4] ZHOU Peng, SHI Wei, TIAN Jun, 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. Berlin, Germany: Association for Computational Linguistics, 2016: 207–212.
- [5] ZHANG Lei, XIANG Fusheng. Relation classification via BiLSTM-CNN[C]//Proceedings of the 3rd International Conference on Data Mining and Big Data. Shanghai, China: Springer, 2018: 373–382.
- [6] XU Yan, MOU Lili, LI Ge, 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: Association for Computational Linguistics, 2015: 1785–1794.
- [7] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long short-Term memory 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: Association for Computational Linguistics, 2015: 1556–1566.
- [8] ZHANG Yuhao, QI Peng, 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: Association for Computational Linguistics, 2018: 2205–2215.
- [9] 甘丽新, 万常选, 刘德喜, 等. 基于句法语义特征的中文实体关系抽取 [J]. 计算机研究与发展, 2016, 53(2): 284–302.
GAN Lixin, WANG Changxuan, LIU Dexi, et al. Chinese named entity relation extraction based on syntactic and semantic features[J]. Journal of computer research and development, 2016, 53(2): 284–302.
- [10] GUO Zhijiang, ZHANG Yan, LU Wei. Attention guided graph convolutional networks for relation extraction [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: ACL, 241–251.
- [11] FU T J, LI P H, MA Weiyun. GraphRel: modeling text as relational graphs for joint entity and relation extraction [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 1409–1418.
- [12] PENG Nanyun, POON H, QUIRK C, et al. Cross-sentence *N*-ary relation extraction with graph LSTMs[J]. Transactions of the association for computational linguistics, 2017, 5: 101–115.
- [13] SONG Linfeng, ZHANG Yue, WANG Zhiguo, 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: Association for Computational Linguistics, 2018: 2226–2235.
- [14] ZHOU Peng, XU Jiaming, QI Zhenyu, et al. Distant supervision for relation extraction with hierarchical selective attention[J]. Neural networks, 2018, 108: 240–247.
- [15] JI Guoliang, LIU Kang, HE Shizhu, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI Press, 2017.
- [16] ZHANG Shu, ZHENG Dequan, HU Xinchun, et al. Bidirectional long short-term memory networks for relation classification[C]//Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation.

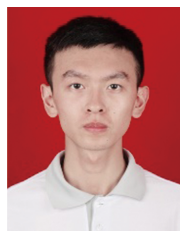
- Shanghai, China: PACLIC, 2015: 73–78.
- [17] ZELENKO D, AONE C, RICHARDELLA A. Kernel methods for relation extraction[J]. The journal of machine learning research, 2003, 3: 1083–1106.
- [18] ZENG Daojian, LIU Kang, LAI Siwei, et al. Relation classification via convolutional deep neural network [C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin, Ireland: Dublin City University and Association for Computational Linguistics, 2014: 2335–2344.
- [19] ZHANG Yuhao, ZHONG V, CHEN Danqi, et al. Position-aware attention and supervised data improve slot filling[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark: Association for Computational Linguistics, 2017: 35–45.
- [20] 马语丹, 赵义, 金婧, 等. 结合实体共现信息与句子语义特征的关系抽取方法 [J]. 中国科学: 信息科学, 2018, 48(11): 1533–1545.
MA Yudan, ZHAO Yi, JIN Jing, et al. Combining entity co-occurrence information and sentence semantic features for relation extraction[J]. Scientia sinica informationis, 2018, 48(11): 1533–1545.
- [21] MINTZ M, BILLS S, SNOW R, et al. Distant supervision for relation extraction without labeled data[C]//Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Suntec, Singapore: Association for Computational Linguistics, 2009: 1003–1011.
- [22] ZENG D, KANG L, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks[C]//Proceedings of the Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processin. Lisbon, Portugal, 2015: 1753–1762.
- [23] HENDRICKX I, KIM S N, KOZAREVA Z, et al. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals[C]//Proceedings of the 5th International Workshop on Semantic Evaluation. Uppsala, Sweden: ACM, 2010: 33–38.
- [24] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: Association for Computational Linguistics, 2014: 1532–1543.
- [25] YU Bowen, ZHANG Zhenyu, LIU Tingwen, et al. Beyond word attention: using segment attention in neural relation extraction[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China: IJCAI, 2019: 33–38.
- [26] 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.

作者简介:



张勇, 教授, 博士生导师, 博士, 中国人工智能学会自然计算与数字智能城市专委会委员, 主要研究方向为智能优化和数据挖掘。主持国家自然科学基金 3 项, 中国博士后科学基金特别资助等省部级科研项目 5 项。获教育部高等学校科学研究优秀成果二等

奖。获授权发明专利 4 项, 发表学术论文 50 余篇。



高大林, 硕士研究生, 主要研究方向为自然语言处理、关系抽取。



巩敦卫, 教授, 博士生导师, 博士, 江苏省自动化学会常务理事、副秘书长, 主要研究方向为智能优化和软件测试。主持国家“973”计划子课题 1 项, 国家重点研发计划子课题 1 项, 国家自然科学基金 6 项, 省部级科研项目 8 项。获高等学校科学研究优秀

成果二等奖、江苏省科学技术二等奖。获授权发明专利 15 项。出版专著 8 部, 发表学术论文 100 余篇。