

基于控制输入长短期记忆网络的关系抽取方法^①



马瑛超, 张晓滨

(西安工程大学 计算机科学学院, 西安 710048)

通信作者: 张晓滨, E-mail: xiaobinzhangcn@126.com

摘要: 目前基于传统深度学习的关系抽取方法在复杂语境下抽取较为困难, 且未考虑语境中非目标关系对关系抽取所带来的影响. 针对这一问题, 本文提出了控制输入长短期记忆网络 CI-LSTM (control input long short-term memory), 该网络在传统 LSTM 的基础上增加了由注意力机制和控制门阀单元组成的输入控制单元, 控制门阀单元可依据控制向量进行关键位置上的重点学习, 注意力机制对单个 LSTM 的输入的不同特征进行计算. 本文通过实验最终选择使用句法依存关系生成控制向量并构建关系抽取模型, 同时使用 SemEval-2010 Task8 关系数据集以及该数据集中具有复杂语境的样本对所提方法进行实验. 结果表明, 相比于传统的关系抽取方法, 本文所提 CI-LSTM 在准确率上有进一步提升, 并在复杂语境中具有更好的表现.

关键词: 关系抽取; 长短期记忆网络; 门控制; 句法依存关系

引用格式: 马瑛超, 张晓滨. 基于控制输入长短期记忆网络的关系抽取方法. 计算机系统应用, 2022, 31(3): 282-287. <http://www.c-s-a.org.cn/1003-3254/8342.html>

Relation Extraction Method Based on Control Input Long Short-term Memory

MA Ying-Chao, ZHANG Xiao-Bin

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: At present, traditional deep learning-based relation extraction methods are unable to extract relations in complex contexts and fail to consider the impact of non-target relations in a context on relation extraction. In response, this study proposes a control input long short-term memory (CI-LSTM) network that adds an input control unit composed of an attention mechanism and a control gate valve unit to the traditional LSTM network. The control gate valve unit can perform focused learning on key positions according to the control vector, and the attention mechanism calculates the different features of the inputs of a single LSTM network. After experiments, this study finally chooses to use syntactic dependency to generate control vectors and build a relation extraction model. An experiment is then conducted on the SemEval-2010 Task8 relation data set and the samples in the data set with complex contexts. The results show that compared with the traditional relation extraction method, the CI-LSTM network proposed in this study achieves further improvement in accuracy and better performance in complex contexts.

Key words: relation extraction; long short-term memory network (LSTM); gate control; syntactic dependency

关系抽取^[1,2]任务是指对实体间的相互关系进行抽取, 其目标是在给定实体对的情况下识别并判断目标实体对之间的特定语义关联. 作为信息抽取^[3]的核心技术之一, 关系抽取能够帮助计算机从自然文本中

识别出实体之间蕴含的语义关系. 知识图谱^[4]作为第三代人工智能的基石, 当下正处于高度发展的阶段, 其构建和应用的过程中都离不开关系抽取的帮助, 并且抽取任务完成的精确度会在一定程度上影响到知识图

① 基金项目: 陕西省自然科学基金 (2019JQ-849); 柯桥纺织产业创新项目 (19KQYB23); 西安工程大学研究生创新基金 (chx2021028)

收稿时间: 2021-04-28; 修改时间: 2021-05-28; 采用时间: 2021-06-08; csa 在线出版时间: 2022-01-24

谱构建的质量及其上层应用的效果优劣. 随着自然语言处理技术的不断发展, 关系抽取的效果也获得了一定的提升, 但面对海量的网络文本, 尤其是丰富多样的语言环境时, 目前关系抽取的效果依旧不能满足当前发展的需求.

1 关系抽取

目前, 可通过以下两类方法来实现关系抽取, 1) 基于机器学习的方法^[5-7]; 2) 基于深度学习的方法. 前者主要指通过由文本能够容易得到的浅层特征, 如词性、句法等特征信息, 从而完成关系抽取的工作; 后者则是利用深度神经网络将文本的浅层特征转换为更高级的语义特征. 相较于第1类方法, 第2类方法能有效地解决人工特征选择、特征提取误差传播两大主要问题.

第2类方法在早期由卷积神经网络 (convolutional neural networks, CNN) 和循环神经网络 (recurrent neural networks, RNN) 为代表. 其中 Zeng 等人^[8]首次借助了 CNN, 用于抽取文本中所包含的词或者句子级别的特征, 最终通过 *Softmax* 层进行关系分类, 从而得到了较好的层级. Xu 等人^[9]将句法特征融入到 CNN 中, 并在数据集 SemEval-2010 Task8 进行实验. 准确率相较于仅使用 CNN 进行抽取的模型提高了 4.1%, 验证了该方法在关系抽取任务上的有效性. 除使用 CNN 作为特征提取网络外, Socher 等人^[10]采用 RNN 用于关系抽取, 实验结果表明 RNN 在该类任务中具有良好的表现. RNN 虽然能很好地学习到文本序列的深层知识, 但相应地也会存在梯度爆炸等弊端.

针对上述弊端, 研究者们通过不断的改进和完善, 提出了 RNN 的变体: 长短期记忆网络 (long short-term memory, LSTM). 其通过“门”机制控制信息的传递, 在 RNN 的基础上解决了长距离依赖以及梯度爆炸问题. Xu 等人^[11]将文本的浅层特征进行融合, 应用 LSTM 完成关系抽取, 实验证明该方法优于传统的 CNN 方法. Zhang 等人^[12]进一步使用 LSTM 的变体双向长短期记忆网络 (Bi-LSTM) 模型进行抽取, 其最佳的实验结果相较于传统方法准确率提高了 14.6%, 可见 Bi-LSTM 模型在关系抽取中能获得更为优异的结果.

随着深度学习的发展, 研究者们试图寻找一种新的结构以帮助模型在学习的过程中对部分特征进行强化学习, 此时注意力机制 (attention)^[13]逐渐步入人们的

视野. 它是由 Treisman 等人^[14]提出的选择算法, 该算法通过注意力机制对特征的关键部分进行重点学习, 对于深度学习网络的学习过程有积极意义. 王红等人^[15]将 attention 引入到 Bi-LSTM 中完成关系抽取, 在实验过程中取得了优秀的成绩. 但基于注意力机制的模型仅通过对输入特征的控制进行计算, 未考虑模型神经元之间特征的相关性. 并且在复杂的语意环境下, 这类模型也容易受到文本中非目标实体对关系语义的干扰, 从而影响方法的性能.

鉴于上述不足, 本文提出了控制输入长短期记忆网络 CI-LSTM (control input long short-term memory), 该网络在传统 LSTM 的基础上增加了由注意力机制和控制门阀单元组成的输入控制单元, 控制门阀单元可依据控制向量进行关键位置上的重点学习, 注意力机制对单个 LSTM 的输入的不同特征与目标实体对信息进行计算, 从而对特征的关键部分进行可控制的选择学习, 有效提升了关系抽取的性能.

2 基于控制输入 LSTM 的关系抽取方法

2.1 输入控制单元

为使网络能够有选择进行学习, 本文受门控机制的启发设计出输入控制单元, 输入控制单元结构如图 1.

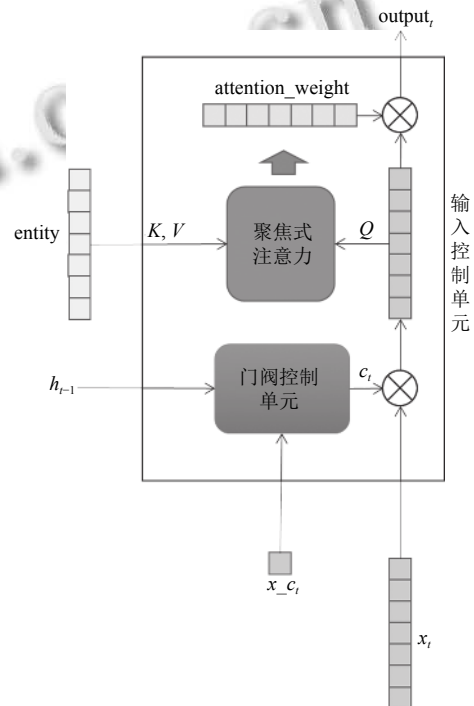


图 1 输入控制单元结构

由图1可以看出,输入控制单元包含两部分.

第一部分为依靠控制向量和门控机制的门阀控制单元,该单元能够根据控制向量对当前时刻的输入进行有选择地进行控制.首先门阀控制单元依据当前时刻的输入和前一时刻的细胞状态计算控制门权重,然后根据控制向量和控制门权重对输入进行计算.

$$c_t = \sigma(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (1)$$

$$x'_t = x_t \cdot c_t^{x-c_t} \quad (2)$$

其中, c_t 为输入控制门的控制信息, x'_t 为更新后的输入信息, x_{-c_t} 为控制向量, W_c 和 b_c 为记忆门的权重及偏置值.

第二部分进行特征级别的注意力机制的计算,本文选择聚焦式注意力机制将经过门阀控制单元计算出的输入与实体信息进行注意力机制的计算,公式如下:

$$a_i = \text{Softmax}(s(x_i, q)) \quad (3)$$

其中, a_i 为第 i 个特征的注意力权重, s 为注意力打分函数, x_i 为 t 时刻第 i 个输入特征, q 为作为 *key* 的实体特征.

在特征级别的注意力计算中,输入为 t 时刻输入特征的值, q 为实体信息的表示,文本在选择注意力打分函数时使用双线性模型,公式如下:

$$s(x_i, q) = x_i W^T q \quad (4)$$

其中, W 为可训练参数,维度与 q 相同.

最终输入控制单元输出的结果 \tilde{x}_t 为:

$$\tilde{x}_t = a \cdot x'_t \quad (5)$$

其中, x'_t 为注意力模块的输入, a 则代表注意力的权重矩阵.

2.2 输入控制 LSTM

在 LSTM 中,最为关键的部分是由细胞状态 C 以及 3 个门(即记忆门 i 、遗忘门 f 与输出门 o) 组成,在时间 t 流入模型的信息由 LSTM 控制单元加以处理,有选择地留下或舍弃 C 的部分信息,并经由计算后传递给下一控制单元.为使得网络能够在学习上下文特征时能够对输入进行有选择地学习,本文在原有 LSTM 结构中引入输入控制单元.输入控制 LSTM 控制单元结构如图2所示.

其中,遗忘门将计算对前一时刻传入的细胞状态进行遗忘的数值,这一部分的计算过程如式(6)所示:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

其中, f_t 为遗忘门的控制信息,用于减少 $t-1$ 时 C 中的

信息, W_f 和 b_f 则分别为遗忘门的权重及偏置值.

记忆门将决定输入的特征有多少能进入新的细胞状态,记忆门的计算过程如式(7)所示:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

其中, i_t 为记忆门的控制信息, W_i 和 b_i 为记忆门的权重及偏置值.

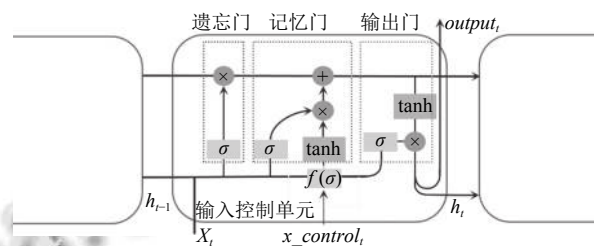


图2 CI-LSTM 细胞结构

输入控制单元依据输入信息和控制向量来实现对输入信息的管理,即对输入的信息有选择地遗忘.输入控制门与记忆门的控制机制不同之处在于输入控制门的控制机制仅作用于输入信息,这种控制方式能够更为直接的作用于输入信息本身,从而在不影响记忆门对细胞状态控制的同时,对输入特征进行有选择地学习.

进一步地,根据遗忘门和记忆门的控制信息对细胞状态进行更新,更新过程如下:

$$C_t = f_t \times C_{t-1} + i_t \times \tanh(W_c \cdot [h_{t-1}, \tilde{x}_t] + b_c) \quad (8)$$

其中, C_t 为新的细胞状态, C_{t-1} 为前一时刻的细胞状态, W_c 、 b_c 分别反映计算临时细胞状态的权重及偏置值.

输出门将计算哪些信息将被输出,这一过程如下所示:

$$o_t = \sigma(W_o \cdot [h_{t-1}, \tilde{x}_t] + b_o) \quad (9)$$

$$h_t = o_t \times \tanh C_t \quad (10)$$

其中, o_t 为该门的控制信息, W_o 、 b_o 则分别反映该门的权重及偏置值.

2.3 基于控制输入 LSTM 的关系抽取模型

本文将关系抽取分为 4 层,分别为预处理层、编码层、学习层以及输出层.

预处理层主要作用在于生成控制向量,并依据 BERT 的需要进行预编码.其中控制向量的生成本文选择依靠句法依存树辅助完成,句法依存树^[16]是文本中句法关系和语法关系的一种表现形式,通过句法依存树能够很清晰地将词与词之间的依存关系表现出来.并且通过句法依存树能够将句子中主干部分进行提炼,

从而弥补语句中表达不清楚的问题。

本文选择以句法依存树所生成的最短依存路径作为控制向量生成的方式。首先本文借助工具生成句子

的句法依存树,并以两实体为两端来寻找最短依存路径,最后基于该路径按原本句子中词的位置生成控制向量。如图3所示。

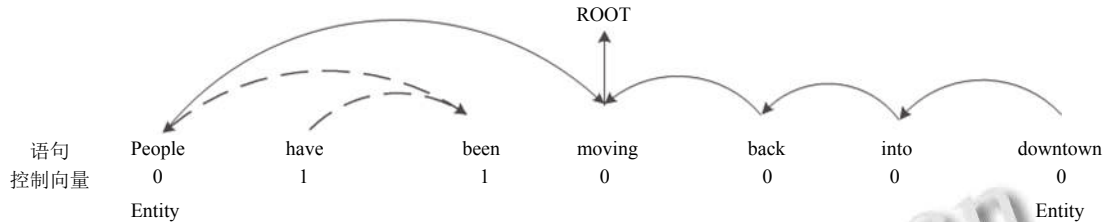


图3 句法依存树的关系表示

控制向量的生成计算如式(11)所示:

$$x_{-c_i} = \begin{cases} 1, & word_i \notin D \\ 0, & word_i \in D \end{cases} \quad (11)$$

其中, x_{-c_i} 为文本中 i 位置的控制向量, $word_i$ 为文本中 i 位置的单词, D 为最短依存路径。

由于自然语言处理中神经网络不能直接输入文本,需要将文本转化为向量的数据。因此编码层使用 BERT^[17] 作为模型的编码层, BERT 模型自推出以来就备受自然语言处理界的青睐,该模型以其优秀的性能已经逐步取代 Word2Vec、ELmo 成为语言模型中不可或缺的一部分。

学习层以双向的控制输入 LSTM 结构进行学习。其中实体信息表示本文通过记录实体开始位置,并从编码层截取实体出现位置的编码向量作为实体信息的表示。同时将首实体与尾实体的拼接作为实体信息的表示输入到控制输入 LSTM 中帮助输入控制单元的注意力计算。

输出层以一个全连接层和 Dropout 层作为中间层, Dropout 层可以在一定程度上减少过拟合的现象,最后通过 Softmax 函数进行判断关系类别的判断。CI-LSTM 模型结构如图4所示。

3 实验分析

3.1 实验数据

为验证所提模型的效果,本文采用 SemEval2010-task8 数据集作为实验语料进行测试。该数据集包含 10 类关系,共 10717 条数据。其中包括 8000 条训练数据和 2717 条测试样本。样本关系分布如表1所示。

除 SemEval2010-task8 数据集外,本文同时选取 150 条 SemEval2010-task8 数据集测试数据中存在多种

语义关系的样本作为复杂语境下的实验样本。

本文评价指标采用准确率 P 、召回率 R 以及 F 值,各指标的具体定义为:

$$P = \frac{T_P}{T_P + F_P} \quad (12)$$

$$R = \frac{T_P}{T_P + F_N} \quad (13)$$

$$F = \frac{2 \times P \times R}{P + R} \quad (14)$$

其中, T_P 表示模型对输出类型判断正确的数据条目; F_P 表示模型对输出关系判断错误的数据条目; F_N 表示模型对输出关系判断错误但实际正确的条目。

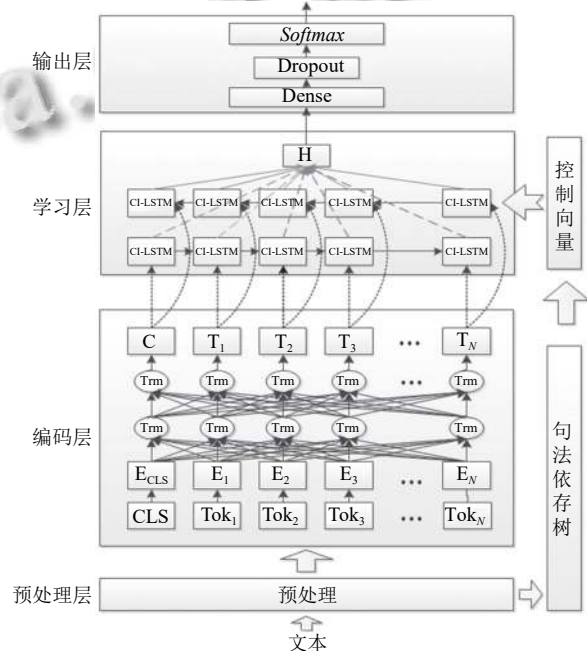


图4 基于 CI-LSTM 的关系抽取模型结构图

表1 SemEval2010-task8 数据集样本关系分布

关系类型	样本个数
其他关系 (others)	1 864
因果关系 (cause-effect)	1 331
整体与部分关系 (component-whole)	1 253
实体与目标关系 (entity-destination)	1 137
实体与来源关系 (entity-origin)	974
生产者与产品关系 (product-producer)	948
成员与集合关系 (member-collection)	923
实体与主题关系 (message-topic)	895
内容与包含关系 (content-container)	732
工具与使用者关系 (instrument-agency)	660

3.2 参数设置

本文实验基于 Keras 深度学习框架, 采用 Adam 优化算法训练模型, 并使用最小化交叉熵的方法进行最优参数的选取. 为方便模型的对比, 在后续试验中, 模型的公共参数均按照最优参数进行实验, 并同样使用 BERT_BASE 作为编码层, 模型的最优参数设置如表 2.

表2 模型最优参数设置

模型参数	值	模型参数	值
初始学习率	0.001	控制向量维度	128
文本最大长度	128	LSTM神经元个数	64
Bert维度	768	批处理样本个数	8
网络迭代次数	4	Dropout系数	0.5

3.3 控制向量生成方式实验

为验证以句法依存树作为控制向量的生成标准的可靠性, 本文以单词相对于实体的位置作为控制向量的生成依据进行了实验. 依据位置的控制向量生成即将文本以实体为界限分为三段式, 在实体之间的信息视作关键信息, 处于两实体两侧的信息视作非关键信息. 以实体位置进行控制向量的生成如图 5 所示.

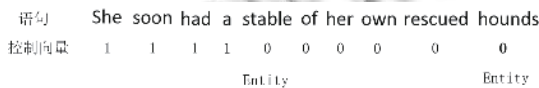


图5 基于实体位置生成控制向量

实验结果如表 3 所示.

表3 实验结果对比 (%)

编码方式	P	R	F
本文方法	73.95	72.18	73.05
基于位置的方法	70.11	69.23	69.67

实验结果表明依据句法依存树的控制向量生成方式在关系抽取中与本文提出的基于输入控制 LSTM 结

构结合能够更好地对关系特征进行学习.

3.4 实验结果

为验证本文模型在关系抽取的效果, 本文分别对本文模型和 RNN、CNN、LSTM、Bi-LSTM、Attention-LSTM 模型进行实验对比, 实验结果如表 4.

表4 SemEval2010-task8 数据集实验结果 (%)

模型	P	R	F
RNN	66.77	43.97	53.02
CNN	67.24	55.24	60.73
LSTM	73.81	65.40	69.35
Bi-LSTM	73.35	66.80	69.92
At-LSTM	73.36	70.29	71.79
CI-LSTM	73.95	72.18	73.05

从表 4 中可以看出, 相较于 RNN、CNN, CI-LSTM 的在 F 上分别提升了 20.03%、12.32%, 这是因为 CI-LSTM 内部采用以 LSTM 的结构来学习语句的特征, 而 LSTM 相对于 CNN 和 RNN 具有更为优秀的结构. 相比于 LSTM 模型和 Bi-LSTM, CI-LSTM 的 F 值分别提升了 3.7% 和 3.13%, 这是因为 CI-LSTM 在传统 LSTM 的基础上增加了控制输入单元, 输入控制单元的引入能够使模型能够对关键部分的特征进行重点学习, 并且针对实体信息的特征级注意力的计算能够帮助模型学习到专注于当前实体对的语义关系信息. CI-LSTM 相较于 Attention-LSTM 效果提升了 1.26%. 这是因为相对于 Attention 机制自动的计算句子各个部分的权重, 在训练过程中容易受到其他关系的干扰, 而 CI-LSTM 可以根据控制向量的值进行选择性的学习, 从而增加了复杂语境下的关系判断精度.

为验证本文模型在复杂语境下的性能, 本文使用 SemEval2010-task8 中选取出的语句中包含多种关系的样本进行单独的测试. 复杂语境实验结果如表 5 所示.

表5 复杂语境模型实验结果 (%)

模型	P	R	F
LSTM	73.81	65.40	69.35
Bi-LSTM	73.35	66.80	69.92
CI-LSTM	72.36	70.29	71.31

这一结果表明本文提出的 CI-LSTM 相较于传统的方法能够在复杂语境下进行更好的关系抽取任务.

4 结论与展望

本文针对当前传统深度学习关系抽取方法在复杂

语境下所存在的文本中其他关系对目标关系抽取产生干扰的问题提出了 CI-LSTM 网络. 该网络在传统 LSTM 的基础上增加了由注意力机制和控制门阀单元组成的输入控制单元, 其中控制门阀单元可依据控制向量进行关键位置上的重点学习, 而注意力机制则对单个 LSTM 的输入的不同特征与实体对信息进行计算. 本文通过句法依存关系进行控制向量的生成, 并以控制输入 LSTM 作为学习层, *Softmax* 函数作为判断关系的方式进行关系抽取模型的构建. 在 SemEval2010-task8 任务数据集上进行实验, 由实验结果可以看出 CI-LSTM 优于其他方法, 能得到更好的关系抽取效果. 另外, 本文通过对 SemEval2010-task8 中具有复杂语境的数据进行了再次的验证, 实验结果表明, 相对于传统深度学习方法, 本文所提方法对复杂语境下的关系抽取更具有有效性. 在后续的研究中, 可以考虑如何在减少参数量的同时根据控制向量进行更好的选择学习, 以及在控制向量生成的方法上进行改进.

参考文献

- 1 刘辉, 江千军, 桂前进, 等. 实体关系抽取技术研究进展综述. 计算机应用研究, 2020, 37(S2): 1-5.
- 2 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述. 软件学报, 2019, 30(6): 1793-1818. [doi: 10.13328/j.cnki.jos.005817]
- 3 郭喜跃, 何婷婷. 信息抽取研究综述. 计算机科学, 2015, 42(2): 14-17, 38. [doi: 10.11896/j.issn.1002-137X.2015.02.003]
- 4 刘岍, 李杨, 段宏, 等. 知识图谱构建技术综述. 计算机研究与发展, 2016, 53(3): 582-600. [doi: 10.7544/issn1000-1239.2016.20148228]
- 5 Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. San Diego: Association for Computational Linguistics, 2004. 22-26.
- 6 Sun X, Dong LH. Feature-based approach to Chinese term relation extraction. 2009 International Conference on Signal Processing Systems. Singapore: IEEE, 2009. 410-414.
- 7 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. 计算机研究与发展, 2016, 53(2): 284-302. [doi: 10.7544/issn1000-1239.2016.20150842]
- 8 Zeng DJ, Liu K, Lai SW, et al. Relation classification via convolutional deep neural network. Proceedings of 25th International Conference on Computational Linguistics. Dublin: ACL, 2014. 2335-2344.
- 9 Xu K, Feng Y, Huang S, et al. Semantic relation classification via convolutional neural networks with simple negative sampling. Computer Science, 2015, 71(7): 941-949.
- 10 Socher R, Huval B, Manning CD, et al. Semantic compositionality through recursive matrix-vector spaces. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Stroudsburg: ACL, 2012. 1201-1211.
- 11 Xu Y, Mou LL, Li G, et al. Classifying relations via long short term memory networks along shortest dependency paths. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015. 1785-1794.
- 12 Zhang S, Zheng DQ, Hu XC, et al. Bidirectional long short-term memory networks for relation classification. 29th Pacific Asia Conference on Language, Information and Computation. Shanghai: ACL, 2015. 73-78.
- 13 邹晓春, 赵歆波, 杨勇佳, 等. 一种基于有监督学习的视觉注意力模型. 中国体视学与图像分析, 2015, 20(3): 201-207.
- 14 Treisman AM, Gelade G. A feature-integration theory of attention. Cognitive Psychology, 1980, 12(1): 97-136. [doi: 10.1016/0010-0285(80)90005-5]
- 15 王红, 史金钊, 张志伟. 基于注意力机制的 LSTM 的语义关系抽取. 计算机应用研究, 2018, 35(5): 1417-1420, 1440. [doi: 10.3969/j.issn.1001-3695.2018.05.029]
- 16 Nivre J. Dependency parsing. Language and Linguistics Compass, 2010, 4(3): 138-152. [doi: 10.1111/j.1749-818X.2010.00187.x]
- 17 Devlin J, Chang MW, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2019. 4171-4186.