

doi: 10.3969/j.issn.0490-6756.2020.02.012

基于多层次动态门控推理网络的文本蕴含识别

张芮, 杨煦晨, 琚生根, 刘宁宁, 谢正文, 王婧妍

(四川大学计算机学院, 成都 610065)

摘要: 现有的文本蕴含模型通常计算一次词级别注意力得到两段文本在不同层面的交互特征,但对于文本不同层面的理解,不同重要词的注意力应该是不同的,并且一次词级注意力推理仅能捕捉到文本对局部特征.针对这个问题,提出一种多层次动态门控推理网络,该网络结合了词级别信息的细粒度推理和句子级别门控机制来动态捕捉文本对的语义信息,并采用不同注意力计算方式提取文本对不同层面的语义特征,共同推理文本对的蕴含关系.本文在两个文本蕴含数据集上均做了实验,相较于基准模型和现有主流模型,准确率提升了0.4%~1.7%,通过消融分析,进一步验证了本文模型各部分结构的有效性.

关键词: 多层次动态推理; sgMatch-LSTM; 注意力机制; 文本蕴含

中图分类号: TP391 **文献标识码:** A **文章编号:** 0490-6756(2020)02-0277-07

Multi-level dynamic gated inference network for recognizing textual entailment

ZHANG Rui, YANG Xu-Chen, JU Sheng-Gen, LIU Ning-Ning, XIE Zheng-Wen, WANG Jing-Yan

(College of Computer Science, Sichuan University, Chengdu 610065, China)

Abstract: Most existing models of recognizing textual entailment (RTE) get the interaction features between a premise and a hypothesis by an attention matrix at word level. However, the attention of important words should be different with the degree of understanding from diverse aspects, and only the local features are captured. To solve the problem above, the model with Multi-level Dynamic Gated Inference Network (MDGIN) is proposed, which combines the fine-grained word-level information and sentence-level gating mechanism to dynamically capture the relationships of text pairs. Moreover, the model extracts the different semantic features by diverse attention ways. In this paper, experiments are carried out on two textual datasets. Compared with the benchmark models and the existing mainstream models, the accuracy is improved by 0.4%~1.7%. The effectiveness of each part of the model is further verified by ablation analysis.

Keywords: Multi-level dynamic inference; sgMatch-LSTM; Attention mechanism; Textual entailment

1 引言

自然语言推理(Nature Language Inference, NLI)任务是自然语言处理的一个重要子任务,能否推理出两段文本之间的关系是机器进一步理解

自然语言的基础.自然语言推理又被称为文本蕴含识别(Recognizing Textual Entailment, RTE),是指给定两段文本 P 和 H ,其中 P 为前提(Premise); H 为假设(Hypothesis),若能从前提 P 的内容中推断出假设 H 的内容,则称 P 和 H 为蕴含关

收稿日期: 2018-08-26

基金项目: 四川省新一代人工智能重大专项(2018GZDZX0039); 四川省科技厅重点研发项目(2018GZ0182)

作者简介: 张芮(1995-),女,硕士生,研究方向为智能信息处理.

通讯作者: 琚生根. E-mail: jsg@scu.edu.cn.

系,否则为非蕴含关系,非蕴含关系可进一步划分为中立关系(P 和 H 的事实无关)和矛盾关系(P 和 H 内容矛盾).文本蕴含技术是关系抽取、机器阅读理解^[1]、对话问答和文本摘要等任务的技术基础,运用文本蕴含技术可进一步提高这些任务的精度.

早期文本蕴含研究通过手工构建大量文本特征^[2-3]来判断文本蕴含关系.而随着深度学习理论的完善和连续语义空间向量^[4]表达能力的提升,基于神经网络的方法及序列优化模型^[5]被广泛用于各个领域^[6-7].目前基于神经网络的文本蕴含研究方法主要有两大类:(1)基于文本编码模型^[8-9];(2)基于交互聚合网络^[10].基于文本编码模型旨在得到一个更优的句子表示,基于交互聚合模型则是构建文本对交互特征用于文本关系分类.现有的交互聚合模型通常基于一次词级注意力得到文本交互特征并用于关系判断,但通过一次词级交互往往不能很好地捕捉到文本对的关系.原因包括:(1)仅通过一次词级注意力捕捉到的文本对信息有限;(2)词级交互重点捕捉局部词关系,当文本对包含多个相同词时,局部相似词对于文本关系的判断影响较大.

基于上述分析,本文作了如下工作:(1)提出句子门控推理结构(Sentence Gated Match LSTM, sgMatch-LSTM),该结构结合了次级别局部交互和句子级别门控机制来动态推理前提和假设文本关系;(2)基于 sgMatch-LSTM 结构提出多层次动态门控推理网络(Multi-level Dynamic Gated Inference Network, MDGIN),该网络基于 sgMatch-LSTM 结构并采用不同注意力交互方式构建文本不同层面的特征,提升了模型对两段文本语义信息的理解.

2 相关工作

早期的文本蕴含研究大多采用特征工程方法,通过相似度特征^[11]、逻辑演算^[12]、基于转换^[13]等方式,构建各种特征来判断文本的蕴含关系.随着斯坦福在 2015 年发布大规模文本蕴含语料(Stanford Nature Language Inference, SNLI)^[14],拉开了基于神经网络的文本蕴含技术新篇章.

基于文本编码的模型是通过训练一个句子级别编码网络,来分别编码前提和假设文本,并将表示向量连接后用于两段文本的关系分类.文本编码模型的编码器则可通过 LSTM、CNN、注意力机制^[15-16]等构造.这种方式可有效提升句子语义表示,但缺少了文本之间的交互,而对于文本蕴含任

务来说,交互特征对蕴含关系的识别十分重要.

基于交互聚合网络则是采用注意力机制对前提和假设文本进行交互. Parikh 等人^[17]提出 De-comp-Attention 模型和 Chen 等人^[10]提出 ESIM 模型均通过计算一次静态词级注意力矩阵来提取前提和假设的交互特征,并将交互特征聚合用于分类.这样存在的问题是如果前一层文本表示稍有偏颇,则会影响局部注意力矩阵的结果,进而得到不准确的交互信息;Rocktäschel 等人^[18],Cheng 等人^[19],Wang 等人^[20],采用动态匹配的思想对前提和假设文本进行迭代推理,每一步的词级交互基于上一步的结果在动态变化,一定程度上缓解了静态词级矩阵的缺陷.上述模型都是基于词级语义进行匹配推理,虽然词级信息可以很好地关注到文本的细节,但是针对某些文本,比如 4.4.2 节中表 4 中的样例 3,前提和假设文本中均包含多个相同词“Angiosperms”,“phylum”,“plants”,导致模型在基于词进行推理时很容易将其判断为蕴含关系,但是实际文本关系为非蕴含.因此,本文在词级别动态推理基础上引入了句子级别语义门控机制用于“修正”词级别推理的局部误差.另外,Wang 等人^[20]模型均只考虑词级信息的一个层面,本文模型通过不同注意力计算方式来提取文本不同层面的特征.

3 模型

本文的模型主要由 3 个部分组成:文本编码层、多层次动态匹配层和输出层.其中动态匹配层分别基于 sgMatch-LSTM 结构(简称 sgM)计算三个不同层面的文本交互特征,对应图 1 中 sgM1, sgM2, sgM3,并将特征聚合用于分类.图 1 显示了 MDGIN 模型整体架构.

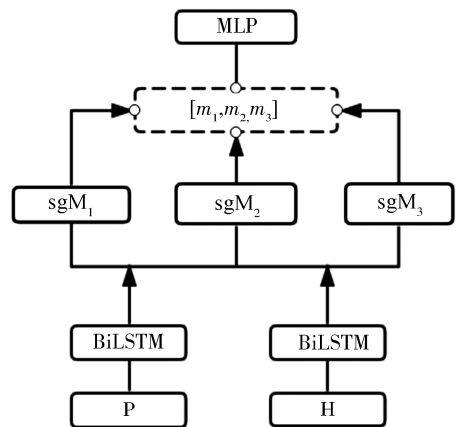


图 1 MDGIN 模型架构图

Fig. 1 Model structure of MDGIN

模型输入形式为 (P, H, Y) , 其中 $P = [p_1, p_2, \dots, p_{l_p}]$, $H = [h_1, h_2, \dots, h_{l_h}]$ 分别为前提文本序列和假设文本序列, l_p 为前提文本长度; l_h 为假设文本长度. Y 为真实的类标签. 表示 p 和 h 之间的蕴含关系, $Y \in \{\text{Entailment}, \text{Neural}, \text{Contradiction}\}$.

3.1 文本编码层

本文使用双向 LSTM (BiLSTM) 网络编码前提和假设的文本向量.

$$\bar{p}_i = \text{BiLSTM}_{enc}(p, i), i \in [1, 2, \dots, l_p] \quad (1)$$

$$\bar{h}_j = \text{BiLSTM}_{enc}(h, j), j \in [1, 2, \dots, l_h] \quad (2)$$

将 BiLSTM 网络前向和后向隐层输出拼接, 得到文本编码向量 \bar{P} 和 \bar{H} , 文本编码向量的输出维度为 $\bar{P} \in \mathbb{R}^{2d \times l_p}$; $\bar{H} \in \mathbb{R}^{2d \times l_h}$; d 表示隐层的维度.

3.2 动态匹配层

3.2.1 Match-LSTM 模型 Wang 等人^[20]提出词级别动态推理模型 Match-LSTM, 本文模型很大程度上受此模型启发.

$$k_{ij} = T(\bar{p}_i, \bar{h}_j, h_{j-1}^m), k_{ij} \in \mathbb{R}^d \quad (3)$$

其中, T 为匹配函数; \bar{p}_i 表示前提文本第 i 个词; \bar{h}_j 表示假设第 j 个词; h_{j-1}^m 表示 $j-1$ 时序隐层向量; k_{ij} 为前提第 i 个词和假设第 j 个词的匹配向量, Match-LSTM 匹配方式如下, 记为 T_1 .

$$T_1 = \tanh(W^p \bar{p}_i + (W^h \bar{h}_j + W^m h_{j-1}^m)) \quad (4)$$

$$\partial_{ij} = \frac{\exp(\omega^k k_{ij})}{\sum_j \exp(\omega^k k_{ij'})} \quad (5)$$

其中, ∂_{ij} 为注意力权重, 表示前提文本第 i 个词和假设文本第 j 个词的对齐权重; $W^p, W^h, W^m \in \mathbb{R}^{2d \times d}$; $\omega^k \in \mathbb{R}^d$ 均为训练参数.

$$\tilde{p}_j = \sum_{i=1}^{l_p} \partial_{ij} \bar{p}_i \quad (6)$$

其中, \tilde{P}_j 表示前提文本关于第 j 个假设词的注意力加权向量. 将加权向量 \tilde{P}_j 和 \bar{h}_j 拼接用于下一个时序的输入 m_{1j} .

$$m_{1j} = [\tilde{p}_j, \bar{h}_j] \quad (7)$$

3.2.2 多层次词级匹配 本文在 Match-LSTM 基础上分别增加两个不同层面的词级注意力 T_2 和 T_3 , 用于提取文本不同层面的特征.

$$T_2 = \tanh(W^p \bar{p}_i * (W^h \bar{h}_j + W^m h_{j-1}^m)) \quad (8)$$

$$T_3 = \tanh(W^p \bar{p}_i - (W^h \bar{h}_j + W^m h_{j-1}^m)) \quad (9)$$

其中, $W^p, W^h, W^m \in \mathbb{R}^{2d \times d}$; $\omega^k \in \mathbb{R}^d$ 均和式(4)相同, 为模型训练参数.

T_2 为元素点乘匹配函数, 点乘注意力用于重点提取两段文本的相似信息, 使模型重点关注文本对相似层面的语义特征. T_3 为元素相减匹配函数, 元素相减操作可以指明推断方向, 同时重点捕捉前提和假设文本之间差异信息. 相比于原 Match-LSTM 模型仅对词进行一次交互, 本文通过多种注意力方式可提取文本不同层面的特征, 提高模型对文本的语义理解.

3.2.3 全局语义门控机制 sgMatch-LSTM 虽然通过多层次词级别推理可充分捕捉文本局部特征, 但由 4.4.2 节表 4 样例 3 可知, 仅从词层面推断, 很容易受到文本中局部语义相似但全局语义相反的情况的影响, 产生错误的判断. 因此本文模型进一步引入句子级别语义作为门控来对词级推理结果进行“修正”. sgMatch-LSTM 结构如图 2 所示. T 为词级匹配函数; S 为全局语义门控机制, 虚线表示 h_{vec} 作用于每步时序.

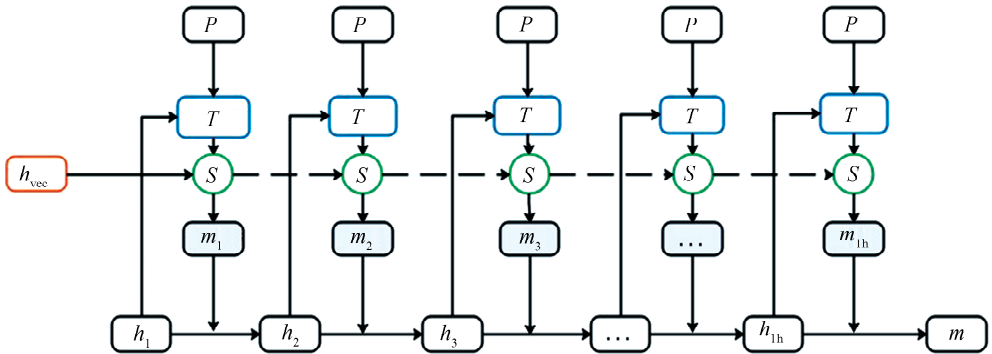


图 2 sgMatch-LSTM 结构
Fig. 2 The structure of sgMatch-LSTM

函数 S 用于计算文本对在句子级别的注意力,并通过 sigmoid 函数映射到 $(0, 1)$ 区间作为门控,用于更新和重新选择词级推理的结果. 如果局部词相似,但全局语义相反,通过句子语义门控选择在一定程度上可降低局部相似词的影响,达到“修正”词级推理的结果. 反之亦然.

$$sg = S(\tilde{p}_j, h_{vec}), sg \in \mathbb{R}^{2d} \quad (10)$$

$$r_j = p_j \odot sg, r_j \in \mathbb{R}^{2d} \quad (11)$$

其中, h_{vec} 为假设文本句向量,通过自注意力^[21]方式得到; sg 为句子级别注意力结果; \odot 为元素相乘; r_j 为门控选择后的前提向量表示. S 为句子级别匹配函数,对于不同层面的句子语义信息采用不同的计算方式,包括 S_1, S_2, S_3 三种计算方式.

$$S_1 = \sigma(W^{pv} \tilde{p}_j + W^{hv} h_{vec} + b_v) \quad (12)$$

其中, $W^{pv}, W^{hv} \in \mathbb{R}^{2d \times 2d}; b_v \in \mathbb{R}^{2d}; \sigma$ 表示 sigmoid 激活函数.

对于句子级别相似和差异层面特征,分别采用点乘和元素相差两种计算方式.

$$S_2 = \sigma(W^{pv} \tilde{p}_j * W^{hv} h_{vec} + b_v) \quad (13)$$

$$S_3 = \sigma(W^{pv} \tilde{p}_j - W^{hv} h_{vec} + b_v) \quad (14)$$

将句子门控修正后的结果 r_j 和 h_j 也分别采用相乘和相差的方式送入下一步时序中,使得相似层面重点关注文本相似特征,差异层面重点关注文本差异特征.

$$m_{2j} = [r_j \odot h_j] \quad (15)$$

$$m_{3j} = [r_j - h_j] \quad (16)$$

3.3 输出层

将 3 个 sgMath-LSTM 最后一步时序的推理向量 m_1, m_2, m_3 拼接得到 m 向量, $m = [m_1, m_2, m_3], m \in \mathbb{R}^{6d}$, 采用一层激活函数为 tanh 的全连接层进行降维,全连接层隐层维度为 \mathbb{R}^{2d} , 最后用 Softmax 函数分类. 模型采用交叉熵损失函数.

4 实验

4.1 数据集和评价指标

本文在 SNLI^[14] 和 SCITAIL 数据集^[22] 进行验证. SNLI 是由斯坦福在 2015 年发布的大型文本蕴含数据集,包含 3 类标签,句式简单而相对固定. 而 SCITAIL 数据集是科学类文本蕴含数据集,根据科学类多选问答任务构造的前提和假设文本,包含 2 类标签,不受手工规则的限定且句式多

样复杂,推理相对困难. 两个数据集分布如表 1 所示. 本任务采用的评价指标:准确率(Accuracy).

表 1 两个数据集的分布

Tab. 1 The distribution of two datasets

N, E, C 分别为 Neural, Entailment, Contradiction

数据集	SNLI	SCITAIL
总样例数	570 000	24 000
训练集	549 367	23 596
验证集	9 842	1 304
测试集	9 824	2 126
类别标签	N, E, C	N, E

4.2 实验环境及参数设置

本文模型基于 Tensorflow 框架搭建,采用 ADAM 优化器^[23] 作为整个模型的优化函数,第一个动量系数设为 0.9; 第二个动量系数设为 0.999. 学习率为 0.0004, batch size 为 32, dropout^[24] 比率为 0.2, l2 正则化在 $[0, 0.0001, 0.00001]$ 之间选择最优. 模型采用预训练的 300 维词嵌入,对于词表外的单词,采用高斯分布随机初始化一个 300 维的向量,所有词向量在整个训练过程中不更新. 本文三个 sgMatch-LSTM 结构共享参数.

4.3 实验结果

本文模型在 SCITAIL 数据集上运行结果如表 2 所示. 由表 2 可知,本文模型在 SCITAIL 数据集取得了 80.7% 的准确率,分别超过 Decomp-Attention^[17] 和 ESIM^[10] 模型 8.4% 和 10.1%. Decomp-Attention 和 ESIM 模型均采用静态词级注意力矩阵进行文本匹配. 但其在较为复杂的 SCITAIL 数据集上表现效果一般,说明仅用一次静态注意力矩阵的方式并不能很好推理复杂的文本关系. 而本文模型采用多层次动态推理方式取得了较好的效果. 本模型分别也超过了 2018 年发布的分解图蕴含模型^[22] (DEGM) 3.4% 以及采用对抗训练的 AdvEntuRE 模型^[25] 1.7%, 验证了本文模型 MDGIN 的有效性.

本文模型在 SNLI 数据集上的验证结果如表 3 所示. DISAN^[15] 完全基于注意力方式对文本进行编码和交互, LSTMN^[19] 引入了记忆网络的思想对 LSTM 进行改进, Word-by-Word Attention^[18] 和 Match-LSTM^[20] 均是动态词级推理模型,但仅采用一次词级推理交互, TreeLSTM^[9] 是基于树结构额外捕捉文本的词法句法信息, DSA^[26] 是 2018 年新发布的基于动态词级注意力的编码模型. 从实

验结果来看, 本文模型均优于上述模型.

表 2 SCITAL 数据集验证结果(准确率)

Tab. 2 Performance (accuracy) on SCITAIL dataset

模型	验证集/%	测试集/%
Majority class	63.3	60.3
Decomp-Attention	75.4	72.3
ESIM	70.5	70.6
Ngram	65.0	70.6
DEGM	79.6	77.3
AdvEntuRE	—	79.0
MDGIN	81.7	80.7

表 3 SNLI 数据集验证结果(准确率)

Tab. 3 Performance (accuracy) on SNLI dataset

模型	训练集/%	测试集/%
word-by-word attention	85.3	83.5
DISAN	91.1	85.6
LSTMN	87.4	85.7
TreeLSTM	93.1	86.0
Match-LSTM	92.0	86.1
Decomp-Attention	89.5	86.3
DSA	87.3	86.8
MDGIN	95.3	87.2

4.4 结果分析

4.4.1 模型结构分析 本节对模型结构进行详细分析, 图 3 中的结构(a)为仅用 LSTM 编码的模型, 结构(b)为词级动态匹配推理模型, 即 Match-LSTM, 结构(c)为增加句子级别门控推理的模型,

结构(d)为多层门控推理网络 MDGIN.

由图 3 可知, 基于词级交互模型(b)相较于仅用 LSTM 编码分别在 SCITAIL 和 SNLI 数据集上准确率提升了 3.9% 和 8.5%, 表明了细粒度文本交互对于文本蕴含任务的重要性. 结构(c)在词级交互基础上增加了句子级别门控机制, 效果进一步提升了 1.1% 和 0.3%, 说明句子级别门控“修正”了词级别匹配结果, 而采用多层次门控推理网络(d), 综合对文本不同层面的推理结果, 准确率进一步提升了 3.6% 和 0.8%, 说明了对文本不同层面采用不同注意力计算方式可有效加强模型对文本特征的提取, 验证了本文模型各部分结构有效性. 另外, 由图 3 可以观察到本文模型对于大型 SNLI 数据集提升较少, 而具有小数据量 SCITAIL 数据集有较大提升, 原因是大数据集本身包含信息足够丰富, 小数据集在信息缺乏情况下基于本文模型可有效捕捉更准确的语义信息和更全面的文本特征, 因此文本关系的判断准确率提升较大.

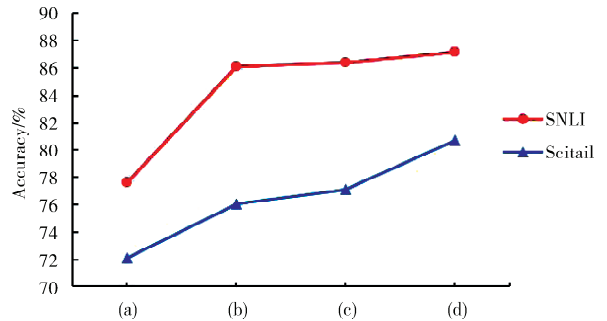


图 3 模型消融折线图

Fig. 3 Ablation performance of MDGIN

表 4 样例结果和分析

Tab. 4 Results and Analysis of examples

样例	文本	Label	MDGIN	Match-LSTM
样例 1	前提 The pupil is the opening through which light enters the eye.	E	E	E
	假设 The pupil of the eye allows light to enter.			
样例 2	前提 This can be dangerous to both plants and animals.	N	N	N
	假设 Nematodes can be a parasite of both.			
样例 3	前提 The angiosperms, or flowering plants, are all members of the phylum	N	N	E
	假设 Angiosperms are the most successful phylum of plants.			
样例 4	前提 Honey bees direct other bees to food sources with the round dance and the waggle dance.	E	E	N
	假设 A bee will sometimes do a dance to tell other bees in the hive where to find food.			

4.4.2 样例分析 表 4 给出了一些文本对样例,由样例 1 和样例 2 可看到,基于词级别的动态推理倾向于对齐前提和假设中相似的词,当相似词越多,越可能被判定为蕴含关系,反之判定为中立关系.而样例 3 中两段文本有较多相似词,仅通过词级别语义推理会将其预测为蕴含关系,但本文模型 MDGIN 引入句子级别语义信息“修正”了局部推理结果.样例 4 由于句子较长且句式相对复杂,Match-LSTM 也错将其预测为中立关系,MDGIN 通过多个层面特征判断可正确预测为蕴含关系.

5 结 论

本文提出 sgMatch-LSTM 结构,通过结合词级别细粒度推理和句子级别门控机制用于动态推理文本蕴含关系,并基于此结构提出多层次动态门控推理网络(MDGIN),分别从三个层面来推理文本的语义关系.本文模型在 SNLI 数据集和较复杂的 SCITAIL 数据集上均有不同程度的提升.由于机器阅读理解、问答等任务也需要蕴含推理技术,未来可尝试将多层次动态推理模型应用于机器阅读理解或对话问答等其他相关领域.

参考文献:

- [1] 王霞,孙界平,琚生根,等.基于段落内部推理和联合问题答案匹配的选择型阅读理解模型[J].四川大学学报:自然科学版,2019,56:423.
- [2] 胡思才,孙界平,琚生根,等.基于扩展的情感词典和卡方模型的中文情感特征选择方法[J].四川大学学报:自然科学版,2019,56:37.
- [3] 金玉,崔兰兰,孙界平,等.基于综合相似度迁移的协同过滤算法[J].四川大学学报:自然科学版,2018,55:477.
- [4] 周顺先,蒋励,林霜巧,等.基于 Word2vector 的文本特征化表示方法[J].重庆邮电大学学报:自然科学版,2018,30:272.
- [5] 刘云,易松.双变换算法在多维序列数据分析中的优化研究[J].四川大学学报:自然科学版,2019,56:633.
- [6] 姚礼垚,熊浩,钟依健,等.基于深度网络模型的牛脸检测算法比较[J].江苏大学学报:自然科:2019,40:197.
- [7] 徐谦,李颖,王刚.基于深度学习的行人和车辆检测[J].吉林大学学报:工学版,2019,49:1661.
- [8] Conneau A, Kiela D, Schwenk H, *et al.* Supervised learning of universal sentence representations from natural language inference data[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark; ACL, 2017.
- [9] Mou L L, Men R, Li G, *et al.* Natural language inference by tree-based convolution and heuristic matching[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany; ACL, 2016.
- [10] Chen Q, Zhu X D, Ling Z H, *et al.* Enhanced LSTM for natural language inference[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada; ACL, 2017.
- [11] Jijkoun V, de Rijke M. Recognizing textual entailment using lexical similarity[C]//Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment. Southampton, UK; Springer, 2005.
- [12] Raina R, Ng A Y, Manning C D. Robust textual inference via learning and abductive reasoning[C]//Proceedings of the 20th national conference on Artificial intelligence-Volume 3. USA; AAAI Press, 2005.
- [13] Bar-Haim R, Dagan I, Greental I, *et al.* Semantic inference at the lexical-syntactic level[C]//Proceedings of the National Conference on Artificial Intelligence. Menlo Park, CA, Cambridge, MA, London; AAAI Press, 1999.
- [14] Bowman S R, Angeli G, Potts C, *et al.* A large annotated corpus for learning natural language inference[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal; ACL, 2015.
- [15] Shen T, Zhou T Y, Long G D, *et al.* Disan: Directional self-attention network for rnn/cnn-free language understanding [C]//Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA; AAAI press, 2018.
- [16] Bahdanau D, Cho K H, Bengio Y. Neural machine translation by jointly learning to align and translate [C]//Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA, USA; arXiv. Org,2015
- [17] Parikh A, Täckström O, Das D, *et al.* A decomposable attention model for natural language inference[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas; ACL, 2016.

- [18] Rocktäschel T, Grefenstette E, Hermann K M, *et al.* Reasoning about entailment with neural attention[C]// Proceedings of the 4th International Conference on Learning Representations. San Juan, Puerto Rico; arXiv. Org, 2016
- [19] Cheng J, Dong L, Lapata M. Long short-term memory-networks for machine reading [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas, USA; ACL, 2016.
- [20] Wang S H, Jiang J. Learning natural language inference with LSTM[C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics. USA; ACL, 2016.
- [21] Lin Z, Feng M, Santos C N, *et al.* A structured self-attentive sentence embedding[C]// Proceedings of the 5th International Conference on Learning Representations. Toulon, France; OpenReview. net, 2017.
- [22] Khot T, Sabharwal A, Clark P. Scitail: A textual entailment dataset from science question answering [C]// Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA; AAAI Press, 2018.
- [23] Kingma D P, Ba J. Adam: a method for stochastic optimization[C]// Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA; arXiv. Org, 2015
- [24] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: a simple way to prevent neural networks from overfitting [J]. J Mach Learn Res, 2014, 15: 1929.
- [25] Kang D, Khot T, Sabharwal A, *et al.* AdvEntuRe: adversarial training for textual entailment with knowledge-guided examples[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Australia; ACL, 2018
- [26] Yoon D, Lee D, Lee S K. Dynamic self-attention: Computing attention over words dynamically for sentence embedding[J]. arXiv preprint arXiv:1808.07383, 2018.

引用本文格式:

中 文: 张芮, 杨煦晨, 琚生根, 等. 基于多层次动态门控推理网络的文本蕴含识别[J]. 四川大学学报: 自然科学版, 2020, 57: 277.

英 文: Zhang R, Yang X C, Ju S G, *et al.* Multi-level dynamic gated inference network for recognizing textual entailment [J]. J Sichuan Univ: Nat Sci Ed, 2020, 57: 277.