

doi: 10.3969/j.issn.0490-6756.2019.03.008

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

王霞¹, 孙界平¹, 琚生根¹, 胡思才^{1,2}

(1. 四川大学计算机学院, 成都 610065; 2. 解放军 61920 部队, 成都 610505)

摘要: 针对当前机器阅读理解方法中仅将问题与段落匹配会导致段落中的信息丢失或将问题和答案连接成单个序列与段落匹配会丢失问题与答案之间的交互, 和传统的循环网络顺序解析文本从而忽略段落内部推理的问题, 提出一种改进段落编码并且将段落与问题和答案共同匹配的模型. 模型首先把段落在多个粒度下切分为块, 编码器利用神经词袋表达将块内词嵌入向量求和, 其次, 将块序列通过前向全连接神经网络扩展到原始序列长度. 然后, 通过两层前向神经网络建模每个单词所在不同粒度的块之间的关系构造门控函数以使模型具有更大的上下文信息同时捕获段落内部推理. 最后, 通过注意力机制将段落表示与问题和答案的交互来选择答案. 在 SemEval-2018 Task 11 任务上的实验结果表明, 本文模型在正确率上超过了相比基线神经网络模型如 Stanford AR 和 GA Reader 提高了 9%~10%, 比最近的模型 SurfaceLR 至少提高了 3%, 超过 TriAN 的单模型 1% 左右. 除此之外, 在 RACE 数据集上的预训练也可以提高模型效果.

关键词: 共同匹配; 多粒度; 机器阅读理解

中图分类号: TP391

文献标识码: A

文章编号: 0490-6756(2019)03-0423-08

Reasoning over intra-document and jointly matching question and candidate answer to the passage based multiple-choice Reading Comprehension

WANG Xia¹, SUN Jie-Ping¹, JU Sheng-Gen¹, HU Si-Cai^{1,2}

(1. College of Computer Science, Sichuan University, Chengdu 610065, China;

2. Troops 61920 of PLA, Chengdu 610505, China)

Abstract: For the current machine reading comprehension method, only matching the question with the paragraph will result in the loss of information of the paragraph or matching the connection of the question and the answer with the paragraph will lose the interaction between the question and the answer. A model that matches an improved encoder of the paragraph with questions and answers is proposed. Firstly, the paragraph is chunked into blocks with multiple granularities, the encoder uses the neural bag-of-words to express the words of each block, and sum the embedding of all words that reside in each block. Next, the blocks are passed into fully-connected layers and expanded to original sequence lengths. The gating function are then constructed through two-layer feed-forward neural network which modeling the relationships between all blocks that each word resides in, allowing for possessing a larger overview of

收稿日期: 2018-11-27

基金项目: 南方电网公司科技资助项目(GZKJXM20170162); 2018 四川省新一代人工智能重大专项(18ZDZX0137)

作者简介: 王霞(1993-), 女, 硕士生, 研究方向为机器智能.

通讯作者: 琚生根. E-mail: jsg@scu.edu.cn

the context information and capturing the intra-document relationships. Finally, the attention mechanism is used to model the interaction between the passage and the question as well as the answer to select an answer. Experimental results on the SemEval-2018 Task 11 demonstrate that our approach's improvement over the baselines such as Stanford AR and GA Reader ranges from 9% to 10%, pulls ahead of recent model SurfaceLR by at least 3% and outperforms the TriAN by 1%. Besides, pretraining the model on RACE datasets helps to improve the overall performance.

Keywords: Joint match; Multi-Granularity; Machine reading comprehension

1 引言

机器阅读问答(Machine Reading for Question Answering, MRQA)研究近年来受到广泛关注,如何让机器能够阅读并且理解自然语言,是认知智能领域研究的主要问题. 阅读理解任务需要在整个文档中整合和推理关于事件,实体及其关系的信息. 问答通常用于评估阅读理解能力^[1]. 机器阅读理解任务的主要类型有完形填空型、选择型和问答型. 深度学习近年来在各领域得到广泛使用,如图像处理^[2,3]等,在机器阅读理解领域也是如此. 相比于基于传统特征的方法,在神经网络中,各种语义单元被表达为连续的语义空间上的向量,可以非常有效地解决语义稀疏性问题^[4]. 各种有效的模型及改进模型或集成模型被研究者提出^[5,6].

目前的阅读理解方法通常基于成对序列匹配,段落与连接问题和候选答案的序列匹配^[7],或段落与问题匹配^[8]. 然而,这些方法可能不适合问题和答案通常同样重要的选择型阅读理解. 仅仅根据问题匹配段落可能没有意义,并且可能导致段落中的信息丢失,例如“Which statement of the following is true?”这样的问题. 另一方面,将问题和答案连接成单个序列进行匹配可能会丢失问题和答案之间的交互信息. 例如,有时答案需要问题来进行指代消解等. 此外,大多数模型通常采用循环神经网络作为编码器,其逐字逐句地顺序解析文本序列,虽然有助于捕获语言学的词法和语法结构,但段落往往较长,这限制了在段落内部进行多句推理.

针对目前研究的不足之处,本文提出的模型,引入了段落内部推理机制,并且将候选答案和问题与段落同时匹配. 首先,类似于自注意力机制^[9],将段落多个粒度下压缩成 NBOW(神经词袋表达). 其次,通过仿射变换重新扩展到原始序列长度. 将得到的多个粒度下的段落表达输入全连接层以形成门控向量应用到段落原始表达上. 该门控建模不同粒度之间关系,捕获段落内部推理,也避免

了每步构造一个门控. 然后,对段落中的每个句子,计算与问题和候选答案的注意力权重向量来构造段落和问题、段落和候选答案的匹配表达. 将两个匹配表达拼接形成共同匹配表达. 最后,对段落的共同匹配表达序列应用层次 LSTM^[10]. 信息从词级别汇总到句子级别,然后从句子级别汇总到文档级别. 这种方式使得模型可以更好地处理需要多句推理的问题.

2 相关工作

大量数据集的提出便于机器阅读理解的研究. 常用的完形填空型数据集 CNN/Dail Maily^[11], Children's Book Test(CBT)^[12]等,其任务是预测问题中缺失的单词或实体;问答型数据集 SQuAD^[13], TriviaQA^[14], NewsQA^[15]等,其任务是给定段落,来回答问题,答案往往是段落中的词或片段;选择性数据集 MCTest^[16], RACE^[17]等.

虽然任务类型不同,但基本框架大致相同,即 1) 将段落、问题、答案(选择型)编码; 2) 建模段落、问题、答案之间的交互; 3) 多步推理; 4) 或预测或生成或选择答案.

Hermann 等人在 CNN/Daily Mail 数据集上,提出使用基于注意力的神经网络模型 Attentive Reader 和 Impatient Reader 来解决阅读理解问题. Attentive Reader 将段落看作单词序列,使用 LSTM 等建模上下文语义信息和词义信息,每个单词的语义由隐藏状态表示;句尾词的隐藏状态表示整个句子的语义信息,然后采用匹配函数来计算段落中每个单词的语义和问题整体语义的匹配程度,再对每个单词的匹配函数值通过 SoftMax 函数进行归一化,最后,相同单词的归一化概率值累加,即为该单词是问题的答案的可能性,选择可能性最大的那个单词作为问题的答案输出. 其中的匹配函数采用前向神经网络的形式. 有很多模型都以此为基础架构,主要不同体现在匹配函数上,如点积^[18]和余弦相似度^[19].

Impatient Reader 与 Attentive Reader 的区别在于它将问题看作单词序列, 使用 LSTM 等建模上下文语义信息和词义信息, 每个单词的语义由隐藏状态表示, 因此, 匹配函数作用在段落中的任意单词和问题中的任意单词之间形成二维矩阵, 然后采用注意力机制作用在行(段落单词和问题中各个单词的语义相似程度)或列(问题单词和文档序列中各个单词的语义相似程度)上. 其变体的区别在注意力机制的设计上^[20,21].

抽取式阅读理解其答案可能在段落的一个句子中, 可以采用信息检索^[22]的方法找到回答问题的支撑语句. 但有时答案不是显而易见的, 需要进行推理, 在神经网络中的推理体现在深层结构上. Hill 等人^[23]在 Children's Book Test 数据集上提出利用记忆网络 (Memory Network, MemNN) 解决阅读理解问题, 它是最早提出推理过程的模型.

从记忆网络的目标函数可以看出, 在输出和回答模块都需要监督, 既需要知道输出选择的相关记忆是否正确, 又需要回答模块生成的答案是否正确. 这限制了模型的推广, 不太容易使用反向传播进行训练. 因此, Perez^[23]提出了一种端到端的记忆网络. Wang^[24]提出的 R-NET 计算段落和问题交互时采用门控注意力机制, 屏蔽掉一些无关的信息, 考虑到段落往往都比较长, 输出层之前的 LSTM 可能无法很好地整合段落的信息, 对段落采用了自注意力机制. Shen^[25]提出的 ReasoNet 采用动态多步推理, 并用强化学习的方法进行训练.

此外, 一些研究工作集中在引入外部知识以及可解释阅读理解方面, 如 Dhingra^[26]在神经网络模型中引入指代消解等语言学信息, Palangi^[27]提出可以用语言学解释的通过优化张量乘积递归网络学习的内部表达.

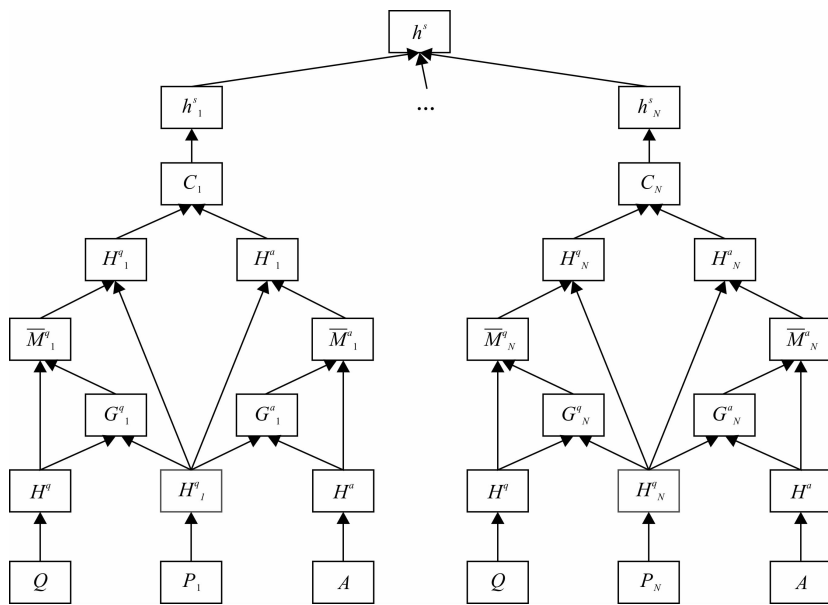


图 1 模型整体框架图
Fig. 1 The whole architecture of the model

3 模型

对于选择型阅读理解的任务, 有一个段落, 一个问题 and 几个候选答案. 其目标是从候选答案中选择正确的答案.

对于每个候选答案构建一个向量, 该向量表示段落与问题和答案的交互. 然后所有候选答案的向量用于选择答案. 本文提出的模型整体框架如图 1 所示.

3.1 输入层

段落, 问题和候选答案对应表达为, $P = \{p_1, p_2, \dots, p_m\}$, $Q = \{q_1, q_2, \dots, q_n\}$ 和 $A = \{a_1, a_2, \dots, a_l\}$, 其中, m, n 和 l 是对应序列的长度. 词嵌入向量: 使用 GloVe^[28] 向量将每个词转换为对应的词嵌入, d 是嵌入维度.

完全匹配特征^[29]: 利用三个简单有效的词汇特征 $EM(P, Q)$, $EM(Q, A)$ 和 $EM(P, A)$ 表明一个中的词 (原始形式和词形还原形式) 是否在另一个中.

3.2 编码层

首先,将词嵌入向量和完全匹配特征输入 Highway Network,分别得到段落、问题和候选答案序列的表达 $H^{p^*} = \{\bar{p}_1, \bar{p}_2, \dots, \bar{p}_m\}$, 其中, $H^{p^*} \in \mathbf{R}^{h \times m}$. $H^q = \{\bar{q}_1, \bar{q}_2, \dots, \bar{q}_n\}$, 其中, $H^q \in \mathbf{R}^{h \times n}$. $H^a = \{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_l\}$, 其中, $H^a \in \mathbf{R}^{h \times l}$. 然后,对段落进行 $k = \{r_1, r_2, \dots, r_k\}$ 次缩放操作,如图 2 所示.

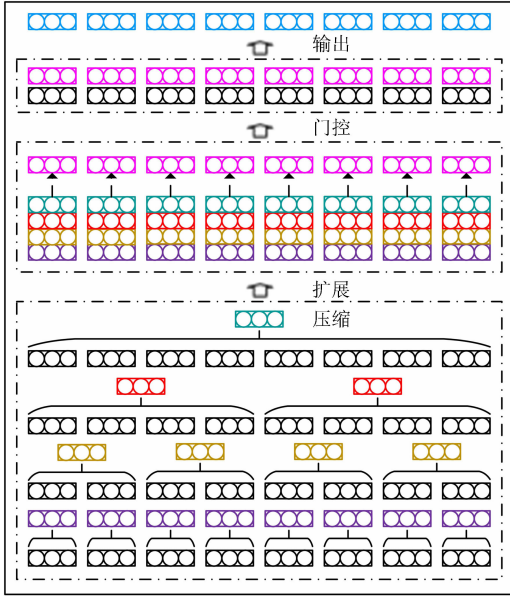


图 2 段落编码

Fig. 2 The encoder of the passage

对每个 r_i , 首先将每隔 r_i 个词的词嵌入相加进行压缩,段落长度变为 m/r_i ,然后将得到的序列输入激活函数为 ReLU 的单层前向神经网络.

$$\bar{p}'_i = \text{ReLU}(W_a \bar{p}_i + b_a) \quad (1)$$

其中, W_a, b_a 是需要学习的参数,且对每个 r_i 参数不共享. 然后,再通过矩阵操作扩展到原来的序列长度.

将得到的 k 个缩放操作得到的序列拼接输入到两层激活函数为 ReLU 的全连接前向神经网络.

$$g_t = F_1(F_2([\bar{p}'_1; \bar{p}'_2; \dots; \bar{p}'_k])) \quad (2)$$

其中, g_t 可以解释为一个从多粒度学习的门控向量来平衡每个词的映射和原始表达. $F_1(\cdot)$ 和 $F_2(\cdot)$ 是激活函数为 ReLU 的单层前向神经网络.

$$z_t = \tanh(W_b \bar{p}_t) + b_b \quad (3)$$

$$y_t = \sigma(g_t) * \bar{p}_t + (1 - \sigma(g_t)) * z_t \quad (4)$$

其中, W_b, b_b 是需要学习的参数, $H^p = \{y_1, y_2, \dots,$

$y_m\}$ 为段落的最终表达, $H^p \in \mathbf{R}^{h \times m}$; σ 是 sigmoid 函数. 这里借鉴了 Highway Network 的思想,不同的是门控函数捕获了多粒度和长期依赖信息.

3.3 交互层

利用注意力机制更新问题和候选答案的表达,使其关注与段落相关的部分. 注意力向量计算如下.

$$G^q = \text{SoftMax}((W^g H^q + b^g \otimes e_Q)^T H^p) \quad (5)$$

$$G^a = \text{SoftMax}((W^g H^a + b^g \otimes e_Q)^T H^p) \quad (6)$$

$$\bar{H}^q = H^q G^q \quad (7)$$

$$\bar{H}^a = H^a G^a \quad (8)$$

其中, $W^g \in \mathbf{R}^{h \times h}$, $b^g \in \mathbf{R}^h$ 是需要学习的参数; $e_Q \in \mathbf{R}^l$ 是元素全为 1 的列向量; $G^q \in \mathbf{R}^{n \times m}$ 和 $G^a \in \mathbf{R}^{l \times m}$ 是赋予问题和候选答案序列隐藏状态的注意力权重; $\bar{H}^q \in \mathbf{R}^{h \times m}$ 是问题隐藏状态的加权和,它表示问题如何与段落的每个隐藏状态对齐; $\bar{H}^a \in \mathbf{R}^{h \times m}$ 同理;将段落表达与问题和候选答案表达共同匹配.

$$M^q = \text{ReLU} \left(W^m \begin{bmatrix} \bar{H}^q \ \$ H^p \\ \bar{H}^q \ominus H^p \end{bmatrix} + b^m \right) \quad (9)$$

$$M^a = \text{ReLU} \left(W^m \begin{bmatrix} \bar{H}^a \ \$ H^p \\ \bar{H}^a \ominus H^p \end{bmatrix} + b^m \right) \quad (10)$$

$$C = \begin{bmatrix} M^q \\ M^a \end{bmatrix} \quad (11)$$

其中, $W^m \in \mathbf{R}^{h \times 2h}$, $b^m \in \mathbf{R}^h$ 是需要学习的参数; C 由两个矩阵按列拼接得到 $\$$ 和 \ominus 是对两个矩阵按元素相减和相乘,其作用是获得更好的局部推理信息^[30]; $M^q \in \mathbf{R}^{h \times m}$ 为该段落表达与注意力加权的问题表达之间的匹配; $M^a \in \mathbf{R}^{h \times m}$ 为段落表达与注意力加权的候选答案表达之间的匹配; $C \in \mathbf{R}^{2h \times m}$ 表示每个段落隐藏状态如何与问题和候选答案匹配.

3.4 层次聚合

为了捕捉段落的句子结构,在共同匹配表达上构建了一个层次 LSTM. 首先将段落分成句子 P_1, P_2, \dots, P_N , 其中 N 是段落中的句子数. 对于每个 $\{P_i, Q, A\}, i \in [1, N]$, 通过输入、编码及交互层可以得到共同匹配表达 C_i . 然后在每个句子的共同匹配表达上应用一个 BiLSTM 层和最大池化层.

$$h_i^s = \text{MaxPooling}(\text{BiLSTM}(C_i))$$

其中, MaxPooling 是按行操作. $h_i^s \in \mathbf{R}^h$ 是共同匹配的句子级表达. 最后,在此基础上再应用一个

BiLSTM层和最大池化层。

$$H^s = [h_1^s; h_2^s; \dots; h_N^s] \quad (12)$$

$$h^t = \text{MaxPooling}(\text{BiLSTM}(H^s)) \quad (13)$$

其中, $H^s \in \mathbf{R}^{h \times N}$ 的句子级表达的按行拼接; h^t 是融合了各种交互信息的最后输出表达。

3.5 输出层

对每个候选答案 A_i , 它最终交互表达为 $h_i^t \in \mathbf{R}^h$. 损失函数计算如下。

$$L(A_i | P, Q) = -\log \frac{\exp(W^T h_i^t)}{\sum_{j=1}^2 \exp(W^T h_j^t)} \quad (14)$$

其中, $W \in \mathbf{R}^h$ 是需要学习的参数。

4 实验

4.1 数据集及参数设置

实验采用 SemEval-2018 Task 11 任务的数据集 MCScript^[31], 并且在 RACE 数据集进行预训

练, RACE 数据集和 MCScript 数据集描述分别如表 1 和表 2 所示. MCScript 数据集由 Mechanical Turk 的工作人员采用第一人称描述一个场景故事, 旨在借用外部常识知识回答问题。

RACE 数据集包含了我国中考和高考中的英文阅读理解问题. 其根据问题的难易程度分为匹配题以及推理题两大类. 匹配题指的是需要读者在段落中找到问题对应的支撑语句, 然后找到问题答案. 推理题则需要读者总结、分析段落中的线索, 进而再选出正确答案. 大约 35% 的题目是匹配题, 65% 是推理题。

在 PyTorch 深度学习框架下实现模型, 使用 300 维的 Glove 向量初始化词嵌入, 且在训练期间不进行微调. k 的取值为 1, 2, 4, 10, 25. 丢失率在 0.1, 0.2 和 0.3 之间调整. 优化器采用 Adam, 且初始学习率为 0.002. 批量大小为 10.

表 1 RACE 数据描述

Tab. 1 RACE data description

Dataset	RACE-M			RACE-M			RACE		
	Train	Dev	Test	Train	Dev	Test	Train	Dev	Test
# passages	6,409	368	362	18,728	1,021	1,045	25,137	1,389	1,407
# questions	25,421	1,436	1,436	62,445	3,451	3,498	87,866	4,887	4,934
Avg. Passage Len		231.1			353.1			321.9	
Avg. Question Len		9.0			10.4			10.0	
Avg. Option Len		3.9			5.8			5.3	
Vocab size		32,811			125,120			136,629	

表 2 MCScript 数据描述

Tab. 2 MCScript data description

Dataset	训练集	验证集	测试集
# passages	1470	219	430
# questions	9731	1411	2797
Avg. Passage Len		196.0	
Avg. Question Len		7.8	
Avg. Option Len		3.6	
Vocab size		100,434	

4.2 对比算法及实验结果

本文模型与基线模型、MCScript 数据集上最近模型以及本文模型进行对比, 结果如表 3 所示。

表 3 对比算法及结果

Tab. 3 Comparison result

模型	验证集(%)	测试集(%)
Sliding Window	—	55
Stanford AR	—	72
GA Reader	—	73
SurfaceLR	81.36	79.66
TriAN-single	83.84	81.94
本文模型	84.81	82.87

(1) Sliding Window. 是一种基于规则的方法来选择与文本具有最高相似度分数的候选答案回答问题. 在段落上采用不同大小的滑动窗口, 计算窗口和答案的词嵌入平均向量的余弦相似度, 具有最高分数的答案被预测为正确答案。

(2) SurfaceLR^[32]. 首先进行词干提取和去除停用词(但不包括‘yes’或‘no’等). 构建段落、问题和答案的长度特征和共现特征, 值得注意的是这个简单模型取得了与神经网络可以竞争的结果。

(3) Stanford Attentive Reader^[33]. 是一种端到端的神经网络方法. 在 CNN/Dail Maily 数据集上取得了很好的效果. 它首次提出问题相关段落表达, 即问题相对段落注意力分数(双线性函数)的加权平均表达. 然后同样采用双线性函数计算问题与答案的相似度, 最后应用 softmax 得到每个答案的预测概率。

(4) Gated-Attention Reader^[34]. 通过注意力机制捕获问题与段落之间的交互, 并采用多层推理迭代地关注文本与问题最相关的部分。

(5) TriAN^[35]. 采用段落与答案、段落与问题、问题与答案三种注意力来建模交互,并且加入词性标注和命名实体识别,基于 ConceptNet 的关系特征,基于 Wikipedia 的词频特征及共现特征. 本文与其单模型进行对比.

在 MCScript 数据集上进行 30 轮训练,在 RACE 数据集上 10 轮迭代. 实验结果如表 3 所示. 可以看出本文模型与基线神经网络模型相比提高了 9%~10%左右,比 SurfaceLR 模型提高了至少 3%,超过 TriAN 单模型 1%左右.

此外,考虑因素:①预训练;②答案匹配;③段落内部推理进行消融实验,实验结果如表 4 所示. 可以看到如果没有在预训练模型正确率降低 1%左右的. 只采用问题进行匹配模型效果降低 4%左右. 在将编码器换作 LSTM 的情况下,模型正确率降低了 2%左右.

表 4 消融实验

Tab. 4 Ablation Study

模型	验证集(%)	测试集(%)
本文模型	84.81	82.87
-预训练	83.68	81.44
-答案匹配	80.81	78.94
-段落编码	82.25	81

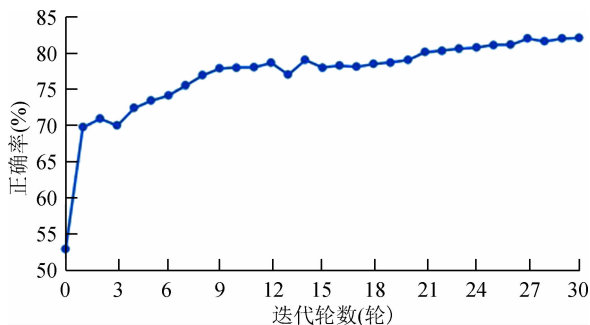


图 3 正确率曲线图

Fig. 3 The accuracy of the model

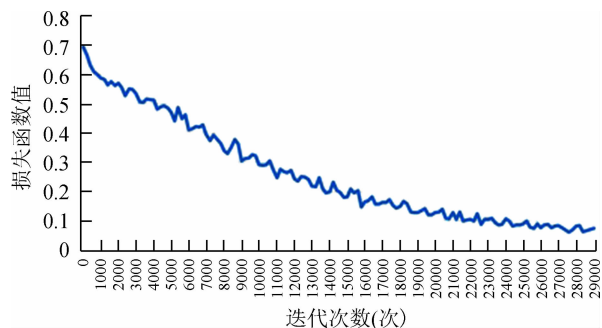


图 4 损失函数曲线图

Fig. 4 The loss of the model

模型训练过程准确率与损失函数曲线如图 3 和图 4 所示. 迭代次数为迭代轮数与训练集数据大小的乘积与批量大小的比值. 图 4 使用每隔 200 迭代次数记录一次的数据制作的曲线图.

5 结论

本文在前人研究的基础上,不同于之前工作只考虑段落与问题匹配,或与问题和候选答案看作一个序列匹配,本文考虑了段落与问题和答案的共同匹配. 同时,针对段落序列往往较长传统神经网络编码无法进行多句推理的问题,利用缩放操作构成的门控机制,捕获多粒度信息,进行段落内部推理,在 SemEval-2018 Task 11 任务上进一步提高了预测准确率. 在后续的工作中考虑问题类型,引入脚本知识,设计可解释性模型,以及对模型集成的研究.

参考文献:

- [1] Kočiský T, Schwarz J, Blunsom P, *et al.* The narrativeqa reading comprehension challenge [J]. *Trans Assoc Comput Linguist*, 2018, 6: 317.
- [2] 吴昊, 平鹏, 孙立博, 等. 基于改进 LRCN 模型的驾驶行为图像序列识别方法 [J]. *江苏大学学报: 自然科学版*, 2018, 39: 303.
- [3] 张雅俊, 高陈强, 李佩, 等. 基于卷积神经网络的人流量统计 [J]. *重庆邮电大学学报: 自然科学版*, 2017, 29: 265.
- [4] 崔一鸣, 刘挺, 王士进. 斯坦福 SQuAD 挑战赛的中国亮丽榜单 [J]. *中国计算机学会通讯*, 2017, 13: 9.
- [5] 高云龙, 左万利, 王英, 等. 基于集成神经网络的短文本分类模型 [J]. *吉林大学学报: 理科版*, 2018, 56: 933.
- [6] 王斯盾, 琚生根, 周刚, 等. 基于集成分类器的用户属性预测研究 [J]. *四川大学学报: 自然科学版*, 2017, 54: 1195.
- [7] Yin W, Ebert S, Schütze H. Attention-based convolutional neural network for machine comprehension [C]//*Proceedings of the Workshop on Human-Computer Question Answering*. Stroudsburg: ACL, 2016.
- [8] Zhu H, Wei F, Qin B, *et al.* Hierarchical attention flow for multiple-choice reading comprehension [C]//*Proceedings of the 30th AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI, 2018.
- [9] Tang D, Qin B, Liu T. Document modeling with

- gated recurrent neural network for sentiment classification [C]//Proceedings of the 2015 conference on empirical methods in natural language processing (EMNLP). Stroudsburg: ACL, 2015.
- [10] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need [C]//Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS). London: MIT, 2017.
- [11] Hermann K M, Kocisky T, Grefenstette E, *et al.* Teaching machines to read and comprehend [C]//Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS). London: MIT, 2015.
- [12] Hill F, Bordes A, Chopra S, *et al.* The goldilocks principle: Reading children's books with explicit memory representations [C]//International Conference on Learning Representations (ICLR). New York: arXiv.org, 2016.
- [13] Rajpurkar P, Zhang J, Lopyrev K, *et al.* Squad: 100,000+ questions for machine comprehension of text [C]//Proceedings of the 2016 conference on empirical methods in natural language processing (EMNLP). Stroudsburg: ACL, 2016.
- [14] Joshi M, Choi E, Weld D S, *et al.* Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Stroudsburg: ACL, 2017.
- [15] Trischler A, Wang T, Yuan X, *et al.* NewsQA: a machine comprehension dataset [C]//Proceedings of the 1st Workshop on Representation Learning for NLP (RepL4NLP). Stroudsburg: ACL, 2016.
- [16] Richardson M, Burges C J C, Renshaw E. Mctest: a challenge dataset for the open-domain machine comprehension of text [C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: ACL, 2013.
- [17] Lai G, Xie Q, Liu H, *et al.* RACE: Large-scale reading comprehension dataset from examinations [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: ACL, 2017.
- [18] Kadlec R, Schmid M, Bajgar O, *et al.* Text understanding with the attention sum reader network [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL). Stroudsburg: ACL, 2016.
- [19] Tseng B H, Shen S S, Lee H Y, *et al.* Towards machine comprehension of spoken content: initial TOEFL listening comprehension test by machine [C]//The Annual Conference of the International Speech Communication Association. Dresden: ISCA, 2016.
- [20] Cui Y, Chen Z, Wei S, *et al.* Attention-over-attention neural networks for reading comprehension [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Stroudsburg: ACL, 2017.
- [21] Wang S, Jiang J. Machine comprehension using match-LSTM and answer pointer [C]//Proceedings of the International Conference on Learning Representations (ICLR). New York: arXiv.org, 2017.
- [22] Clark P, Etzioni O, Khot T, *et al.* Combining retrieval, statistics, and inference to answer elementary science questions [C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2016.
- [23] Liu F, Perez J. Gated end-to-end memory networks [C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL). Stroudsburg: ACL, 2017.
- [24] Wang W, Yang N, Wei F, *et al.* Gated self-matching networks for reading comprehension and question answering [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Stroudsburg: ACL, 2017.
- [25] Shen Y, Huang P S, Gao J, *et al.* ReasonNet: learning to stop reading in machine comprehension [C]//Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017.
- [26] Dhingra B, Jin Q, Yang Z, *et al.* Neural models for reasoning over multiple mentions using coreference [C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL). Stroudsburg: ACL, 2018.
- [27] Palangi H, Smolensky P, He X, *et al.* Question-answering with grammatically-interpretable representations [C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2018.
- [28] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation [C]//Proceedings of the Conference on Empirical Methods in Natural

- Language Processing (EMNLP). Stroudsburg: ACL, 2014.
- [29] Chen D, Fisch A, Weston J, *et al.* Reading wikipedia to answer open-domain questions [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(ACL). Stroudsburg: ACL, 2017.
- [30] Chen Q, Zhu X, Ling Z, *et al.* Enhanced LSTM for natural language inference [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2017.
- [31] Ostermann S, Modi A, Roth M, *et al.* MCScript: A novel dataset for assessing machine comprehension using script knowledge [C]//Proceedings of the 11 International conference on Language Resources and Evaluation. Stroudsburg: ACL, 2018.
- [32] Merkhofer E, Henderson J, Bloom D, *et al.* MITRE at SemEval-2018 Task 11: Commonsense Reasoning without Commonsense Knowledge [C]// Proceedings of the 12th International Workshop on Semantic Evaluations. Stroudsburg: ACL, 2018.
- [33] Chen D, Bolton J, Manning C D. A thorough examination of the cnn/daily mail reading comprehension task [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2016.
- [34] Dhingra B, Liu H, Yang Z, *et al.* Gated-attention readers for text comprehension [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Stroudsburg: ACL, 2017.
- [35] Wang L. Yuanfudao at SemEval-2018 Task 11: three-way attention and relational knowledge for commonsense machine comprehension [C]//Proceedings of the 12th International Workshop on Semantic Evaluations. Stroudsburg: ACL, 2018.

引用本文格式:

中文: 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.

英文: Wang X, Sun J P, Ju S G, *et al.* Reasoning over intra-document and jointly matching question and candidate answer to the passage based multiple-choice Reading Comprehension [J]. J Sichuan Univ; Nat Sci Ed, 2019, 56: 423.