



## 基于时间卷积网络的多项选择机器阅读理解

杨姗姗, 姜丽芬, 孙华志, 马春梅

(天津师范大学 计算机与信息工程学院, 天津 300387)

**摘 要:** 机器阅读理解是自然语言处理领域中一项具有挑战性的任务, 其旨在回答与文章相关的问题, 且需要复杂的语义推理。针对现有机器阅读理解方法提取特征时存在一定程度的信息丢失, 且无法捕获全局的语义关系等问题, 在时间卷积网络(TCN)的基础上, 构建一种多项选择机器阅读理解 M-TCN 模型。采用注意力机制对文章、问题和候选答案进行匹配, 并建立三者之间的内在联系。同时, 为提取高层特征以减少信息丢失, 利用 TCN 对匹配表示进行聚合。通过在公开阅读理解 RCAE 数据集上验证模型的性能, 实验结果表明, 与现有机器阅读理解模型 ElimiNet、MRU、HCM 等相比, 该模型对正确答案的预测精度达到了 52.5%, 且综合性能更优。  
**关键词:** 机器阅读理解; 多项选择; 时间卷积网络; 注意力机制; 深度学习

开放科学(资源服务)标志码(OSID):



中文引用格式: 杨姗姗, 姜丽芬, 孙华志, 等. 基于时间卷积网络的多项选择机器阅读理解[J]. 计算机工程, 2020, 46(11): 97-103.

英文引用格式: YANG Shanshan, JIANG Lifan, SUN Huazhi, et al. Multiple choice machine reading comprehension based on temporal convolutional network[J]. Computer Engineering, 2020, 46(11): 97-103.

## Multiple Choice Machine Reading Comprehension Based on Temporal Convolutional Network

YANG Shanshan, JIANG Lifan, SUN Huazhi, MA Chunmei

(College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China)

**【Abstract】** As a challenging task in the field of natural language processing, machine reading comprehension aims to answer questions related to articles and requires complex semantic reasoning. To solve the problem of information loss and inability to capture the global semantic relationship in feature extraction of existing machine reading comprehension methods, this paper constructs a multiple choice machine reading comprehension M-TCN model based on Temporal Convolutional Network(TCN). Attention mechanism is used to match articles, questions and candidate answers, and the internal relationship among them is established. At the same time, in order to extract high-level features to reduce information loss, TCN is used to aggregate the matching representation. The performance of the model is verified on the RCAE dataset of public reading comprehension. The experimental results show that compared with the existing machine reading comprehension models including ElimiNet, MRU and HCM, the proposed model increases the prediction accuracy to 52.5%, and its comprehensive performance is better.

**【Key words】** machine reading comprehension; multiple choice; Temporal Convolutional Network(TCN); attention mechanism; deep learning

DOI:10.19678/j.issn.1000-3428.0055628

### 0 概述

机器阅读理解是在限定上下文的条件下为给定问题找到答案的一类任务, 它是自然语言处理的核心任务之一<sup>[1]</sup>。近年来, 机器阅读理解已经成为

评价人工智能系统在自然语言处理领域的一个重要应用问题, 计算语言学领域对其给予了极大的关注。传统的机器阅读理解系统主要依赖于语言注释、结构化的外界知识和语义分析<sup>[2]</sup>, 对专业知识的要求较高, 不具备普遍适用性。随着深度学习

基金项目: 天津市自然科学基金(18JCQNJC70200, 18JCYBJC85900)。

作者简介: 杨姗姗(1994—), 女, 硕士研究生, 主研方向为自然语言处理; 姜丽芬、孙华志, 教授、博士; 马春梅, 讲师、博士。

收稿日期: 2019-07-31 修回日期: 2019-10-24 E-mail: wxxjlf@sina.com

在自然语言处理领域的广泛应用,基于深度神经网络的机器阅读理解也迅速发展<sup>[3-5]</sup>。当前机器阅读理解一般分为完形填空式、问答式和多项选择式三类。

早期基于深度学习的机器阅读理解主要是完形填空式。完形填空式任务是在机器阅读并理解一篇文章的内容后,要求机器对指定问题进行回答,而问题一般是被抽掉某个实体词的句子,机器回答问题的过程就是预测问题句子中被抽掉的实体词,且一般要求被抽掉的实体词在文章中出现过。2015 年,大型的完形填空阅读理解数据集 CNN/DailyMail<sup>[2]</sup> 和 CBT<sup>[6]</sup> 的相继公布为机器阅读理解提供了丰富的语料,基于深度学习的机器阅读理解方法也随之得到广泛应用。文献[7]将注意力机制应用于机器阅读理解,并提出了经典的 ASReader 模型。文献[8]基于多跳结构提出门控注意力模型。文献[9]提出用于机器阅读理解的迭代交替注意力机制。但是有学者对 CNN/DailyMail 进行详细评估后发现,该类任务需要处理的逻辑相对简单,现有的针对完形填空式任务的机器阅读理解方法基本已经达到了该类任务的准确度上限<sup>[10]</sup>。

由于完形填空式任务的解决相对简单,学者们试图构造出更加复杂的阅读理解数据集来增强机器对文本的推理,而近年来问答式阅读理解数据集 SQuAD<sup>[11]</sup> 和 NewsQA<sup>[12]</sup> 引起了广泛关注,问答式机器阅读理解也随之兴起。问答式任务是在给定文章和问题的条件下,机器阅读进行一定的推理,进而预测出正确答案,且正确答案一般是文章中的文本段。目前,很多有效的机器阅读理解模型已经被应用于问答式任务,文献[13]使用双向注意力机制对文章和问题进行匹配,并构建了 BiDAF 模型。文献[14]将自匹配注意力机制应用于序列的自关注,提出自匹配网络。文献[15]将卷积神经网络、多头注意力机制和前馈神经网络相结合作为编码器,并提出 QANet 模型,但是由于问答式任务中正确答案是文章中的文本段,一般通过检索出答案的开始位置和结束位置即可得到答案,因此多数面向问答式任务的模型主要关注浅层的语义,并不适用于更深层次的推理。

当前机器阅读理解任务面临的主要挑战仍然是多句推理,这对挖掘深层语义提出更高的要求。针对多项选择式和问答式任务的机器阅读理解方

法存在的上述局限性,且不能很好地解决该问题,因此进行复杂语义推理的多项选择式任务逐渐占据主流。多项选择式任务是给定了文章、问题和候选答案,机器经过理解和推理,从候选答案中选择出正确的答案,候选答案多数是由句子组成,且一般不会在原文中出现。相比于完形填空式和问答式任务,由于正确答案多数无法在原文中找到,多项选择式任务将会更加复杂、更具有代表性,因此本文以多项选择式任务为研究内容。面向多项选择的机器阅读理解模型主要是基于循环神经网络,文献[16]使用双向 GRU 进行特征提取,并结合注意力机制提出分层注意力流模型,文献[17]使用分层的 双向长短期记忆(Long Short-Term Memory, LSTM)网络分别提取句子级别和文档级别的特征。利用循环神经网络提取特征,尽管从理论上能够捕获无限长的记忆,从而建立上下文间的长期依赖关系,但是仍存在梯度消失问题,且随着需要处理序列的增长,还有一定程度的信息丢失,导致无法捕获全局的语义关系。

为有效地对文章、问题和候选答案进行匹配以减少特征的丢失,基于时间卷积网络(Temporal Convolutional Network, TCN)<sup>[18]</sup>,本文构建一种面向多项选择的机器阅读理解 M-TCN 模型。该模型在中国初高中英语考试的阅读理解数据集 RACE<sup>[19]</sup> 上进行训练,并与现有机理阅读理解模型进行对比分析,以验证该模型的机器阅读理解性能。

## 1 模型搭建

针对多项选择式任务,本文采用深度学习技术构建一种 M-TCN 模型,该模型结构如图 1 所示。在图 1 中,假设  $P$  表示文章,  $Q$  表示问题,  $A$  表示一个候选答案,该模型通过预训练好的词向量文件将  $P$ 、 $Q$  和  $A$  中的每个单词向量化,并采用单层的双向 LSTM 网络<sup>[20]</sup> 对向量化后的序列进行编码,以获得文章、问题和候选答案的上下文表示。为了对文章、问题和候选答案进行匹配,构建三者之间的内在联系,分别计算文章和问题以及文章和候选答案之间的注意力,并利用融合函数将注意力值与文章的上下文表示相融合。同时,使用 TCN 和最大池化方法对匹配表示分层进行句级聚合和文档级聚合。M-TCN 模型的目标是计算出每一个候选答案的概率,并以概率最大的候选答案作为预测标签。

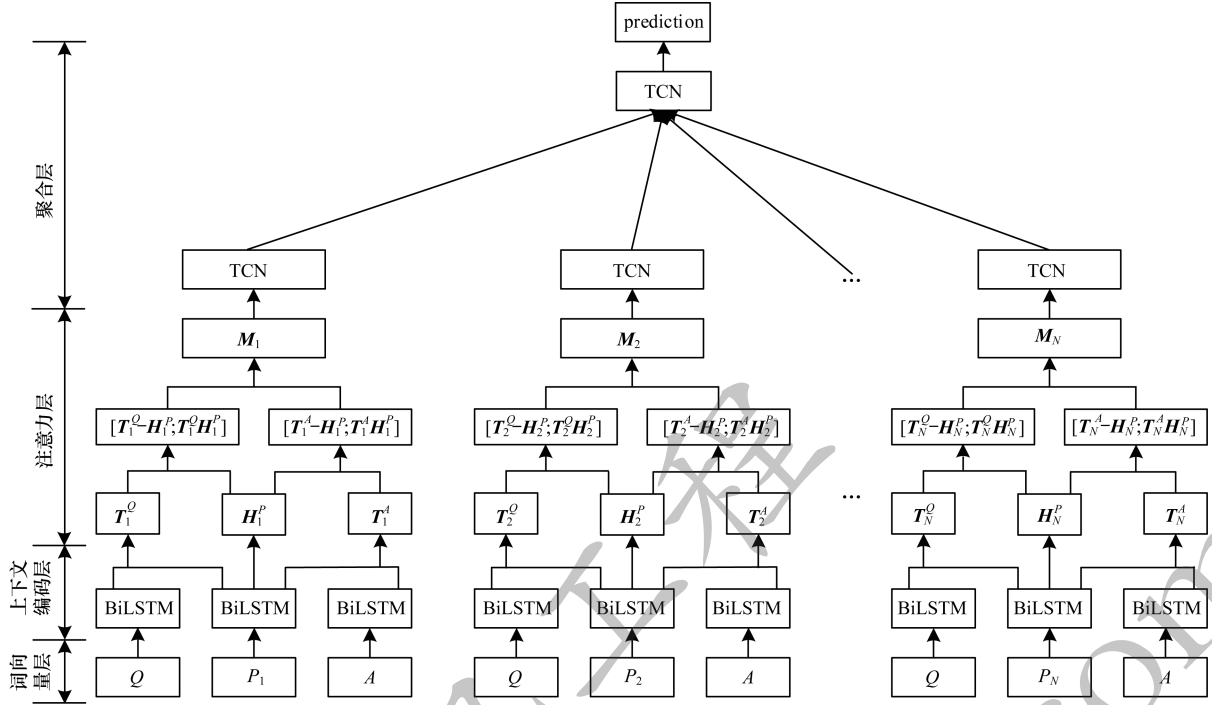


图 1 本文模型结构

Fig. 1 Structure of the proposed model

### 1.1 词向量层

词向量层的主要功能是将文章、问题和候选答案中的每一个单词映射到一个高维向量空间,对于词向量的构建,选择已经预训练好的 300 维 Glove 词向量文件来映射,并将映射后的文章、问题和候选答案分别表示为  $E^p \in \mathbb{R}^{|P| \times 300}$ ,  $E^q \in \mathbb{R}^{|Q| \times 300}$ ,  $E^a \in \mathbb{R}^{|A| \times 300}$ ,其中,  $|P|$ 、 $|Q|$  与  $|A|$  分别表示文章、问题以及候选答案的长度。

### 1.2 上下文编码层

为了利用周围单词的信息来细化单词的表示,本文在上一层提供的向量上,采用 LSTM 对单词之间的相互作用关系进行建模,利用双向 LSTM 将 2 个 LSTM 的输出拼接起来,可由文章词向量  $E^p$  得到文章上下文表示  $H^p \in \mathbb{R}^{|P| \times 21}$ ,由问题词向量  $E^q$  得到问题上下文表示  $H^q \in \mathbb{R}^{|Q| \times 21}$ ,由候选答案词向量  $E^a$  得到候选答案上下文表示  $H^a \in \mathbb{R}^{|A| \times 21}$ ,因为进行拼接的正向 LSTM 和反向 LSTM 的输出都是 1 维的,所以矩阵  $H^p$ 、 $H^q$ 、 $H^a$  都是 21 维的。

### 1.3 注意力层

注意力层主要负责匹配和融合来自文章、问题和候选答案的信息。本文使用注意力机制同时对文章和问题以及文章和候选答案进行匹配,在存在监督训练的过程中,通过更新参数矩阵学习来显示出问题和候选答案中与文章密切相关的内容,忽略与

文章不相关的内容,注意力向量的计算方法如下:

$$S^q = \text{Softmax} \left( H^p \delta \left( (H^q W^m + b^m)^T \right) \right)$$

$$S^a = \text{Softmax} \left( H^p \delta \left( (H^a W^m + b^m)^T \right) \right)$$

$$T^q = S^q H^q$$

$$T^a = S^a H^a \quad (1)$$

其中,  $S^q \in \mathbb{R}^{|P| \times |Q|}$  是文章和问题之间的注意力权重矩阵,  $S^a \in \mathbb{R}^{|P| \times |A|}$  是文章和候选答案之间的注意力权重矩阵,  $W^m$  和  $b^m$  是需要学习的参数,  $\delta$  表示激活函数,本文选择的函数为 ReLU 激活函数。  $T^q \in \mathbb{R}^{|P| \times 21}$  为问题所有隐藏状态的加权和,表示关注文章的问题向量,  $T^a \in \mathbb{R}^{|P| \times 21}$  为候选答案所有隐藏状态的加权和,表示关注文章的候选答案向量。

为了整合原始上下文表示,从不同层次反映语义信息,本文将文章上下文表示  $H^p$  分别与问题注意力向量  $T^q$  以及候选答案注意力向量  $T^a$  相融合,并产生矩阵  $Q'$  和  $A'$ ,  $Q'$  和  $A'$  的计算方法为:

$$Q' = \beta(H^p, T^q)$$

$$A' = \beta(H^p, T^a)$$

其中,  $\beta$  是一个需要训练的融合函数,最简单的融合方法是将 2 个输入相结合,然后进行线性或非线性变换。为了更好地提取融合特征,文献[21]提出一种基于差异性与相似性的融合方法,该方法能有效组合不同的表示形式,同时本文参考文献[17]采用的融合方式对  $\beta$  进行如下定义:

$$\beta(\mathbf{H}^p, \mathbf{T}^q) = \delta([\mathbf{T}^q - \mathbf{H}^p; \mathbf{T}^q \cdot \mathbf{H}^p] \mathbf{W}^z + \mathbf{b}^z) \quad (3)$$

其中,  $[\cdot]$  表示矩阵的连结操作, “ $-$ ” “ $\cdot$ ” 表示 2 个矩阵之间逐元素的减法和乘法操作,  $\mathbf{W}^z$  和  $\mathbf{b}^z$  是需要学习的参数。

将  $\mathbf{Q}'$  和  $\mathbf{A}'$  相结合得到注意力层的输出, 计算方法为:

$$\mathbf{M} = [\mathbf{Q}'; \mathbf{A}'] \quad (4)$$

其中,  $\mathbf{M} \in \mathbb{R}^{|P| \times 4l}$  是文章、问题和候选答案之间的匹配表示。

#### 1.4 聚合层

为了捕获文章的句子结构, 建立句子之间的联系, 本文在匹配层的基础上建立了聚合层。首先将文章分为单个的句子, 并用  $P_1, P_2, \dots, P_N$  分别表示文章中的每个句子, 其中,  $N$  表示文章中句子的总个数。对于每一个三元组  $(P_n, Q, A)$ ,  $n \in [1, N]$ , 通过前几层的处理可以得到匹配表示  $\mathbf{M}_n$ 。为了捕获  $\mathbf{M}_n$  中上下文间的长距离依赖关系, 减少特征丢失, 本文采用 TCN 来提取高层特征。

TCN 是一种特殊的一维时间卷积网络, 用于序列建模任务, 文献 [15] 表明, TCN 在多种任务和数据集上的性能均优于 LSTM 等典型的循环神经网络, 且相比于循环神经网络, TCN 可以记住更长的历史信息, 捕获更全局的语义关系。在序列建模任务中, 一般有以下 2 个要求:

- 1) 模型产生与输入长度相同的输出。
- 2) 信息的传递是单向的, 且未来的信息仅由过去的信息决定。

对于一个一维的输入序列  $\mathbf{I} \in \mathbb{R}^s$  和一个卷积核  $f: \{0, 1, \dots, k-1\} \rightarrow \mathbb{R}$ , 在序列元素  $x$  上的空洞卷积操作  $F$  定义为:

$$F(x) = (\mathbf{I}_d \times f)(x) = \sum_{j=0}^{k-1} f(j) \cdot \mathbf{I}_{x-d, j} \quad (5)$$

其中,  $d$  表示空洞因子,  $k$  表示卷积核的大小, 空洞卷积如图 2 所示。TCN 中还加入了残差连接<sup>[22]</sup>、权值归一化<sup>[23]</sup>和空间 Dropout<sup>[24]</sup>来提高性能, TCN 的整体架构如图 3 所示。

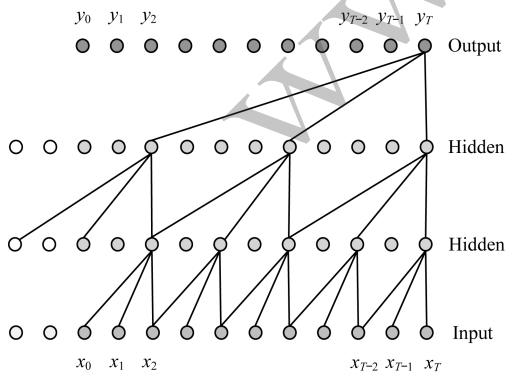


图 2 空洞卷积示意图

Fig. 2 Schematic diagram of hole convolution

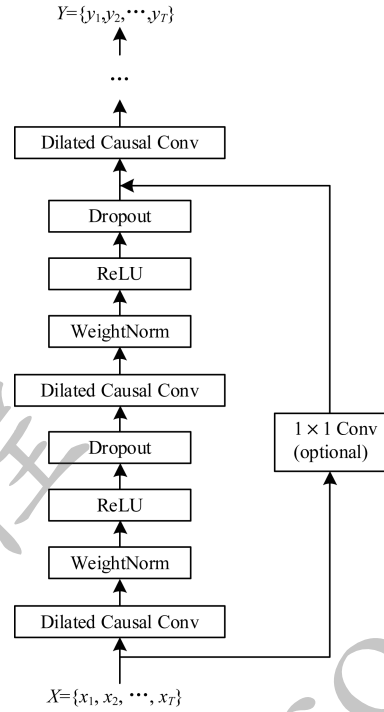


图 3 TCN 整体结构

Fig. 3 The overall structure of TCN

本文模型采用 2 个分层的 TCN, 第一层 TCN 的输入是匹配表示  $\mathbf{M}_n$ , 得到的输出紧接着进行最大池化, 计算方法为:

$$\mathbf{S}_n = \text{MaxPooling}(\text{TCN}(\mathbf{M}_n)) \quad (6)$$

其中,  $\mathbf{S}_n \in \mathbb{R}^d$  是匹配表示  $\mathbf{M}_n$  的句级聚合,  $d$  是 TCN 输出层的维度。所有的句级聚合表示将被另一个 TCN 进一步聚合, 且计算方法为:

$$\mathbf{S} = [\mathbf{S}_1; \mathbf{S}_2; \dots; \mathbf{S}_N] \\ \mathbf{D} = \text{MaxPooling}(\text{TCN}(\mathbf{S})) \quad (7)$$

其中,  $\mathbf{S} \in \mathbb{R}^{N \times d}$  是所有句级聚合表示的串联, 也是更高级别 TCN 的输入,  $\mathbf{D} \in \mathbb{R}^d$  是文章、问题和候选答案之间的最终匹配结果。

对于每一个候选答案  $A_i$ , 本文模型都可以计算出其关于文章和问题的匹配结果  $\mathbf{D}_i \in \mathbb{R}^d$ , 模型损失函数的计算方法为:

$$L(A_i | P, Q) = -\ln \frac{\exp(\mathbf{W}^v \mathbf{D}_i)}{\sum_{j=1}^4 \exp(\mathbf{W}^v \mathbf{D}_j)} \quad (8)$$

其中,  $\mathbf{W}^v$  是需要学习的参数。

## 2 实验与性能分析

### 2.1 数据集

机器阅读理解的效果与采用的数据集密切相关, 本文采用 RACE 数据集。RACE 数据集是一个基于中国初高中英语考试的多项选择阅读理解数据

集,由 27 933 篇文章和 97 687 个问题组成,每个问题配有 4 个候选答案,且只有一个正确答案。为了对应初中和高中阅读理解难度水平,该数据集包含 2 个子数据集,分别为初中阅读理解数据集 RACE-M 和高中阅读理解数据集 RACE-H。将 RACE 数据集看成三元组  $\{P, Q, A\}$  的集合。实验数据的具体信息如表 1 所示,该数据集的样例如表 2 所示。

表 1 实验数据信息

Table 1 Experimental data information

数据集	字集	文章数	问题数
RACE-M	Train	6 409	25 421
	Dev	368	1 436
	Test	362	1 436
RACE-H	Train	18 782	62 445
	Dev	1 021	3 451
	Test	1 045	3 498
RACE	Train	25 137	87 866
	Dev	1 389	4 887
	Test	1 407	4 934
	All	27 933	97 687

表 2 RACE 数据集的样例

Table 2 Example of RACE dataset

名称	样例
文章	Welcome every morning with a smile. Let your first hour set the theme of success and positive action that is certain to echo through your entire day. Today will never happen again. Don't waste it with a false start or no start at all. You were not born to fail. Starting the day right can give good momentum for the rest of the day...Here are some tips on getting the most out of the morning. Wake up early. In theory, there's no difference in waking up early or late as long as you get the proper amount of sleep...Our to-do list might be never ending, but put on top of the list a couple of urgent tasks that will define the day to be a successful one. Check your emails. Only once. Getting a lot of emails can be very distractive. You actually are more productive when you check them only a few times at fixed time of the day.
问题	The purpose for the author to write this passage is to _ A. Tell you to get up early every day B. Offer you the best way to start the day C. Give you advice about how to arrange a day D. Advise you to prioritize your to-do list every morning
答案	B

## 2.2 实验环境和参数设置

本文模型在配备有英特尔 i7 CPU 以及 NVIDIA GTX1080Ti GPU 的计算机上进行训练,采用深度学习框架 Pytorch,集成开发环境为 Pycharm。在深度学习中模型参数的调整对实验结果有重要的影响,参数过大或过小都会影响模型的精度。本文通过实验并结合经验手动调节参数,对模型进行科学训练,

以提高资源利用率。其中,参数的调整包括 TCN 层数、TCN 隐层节点数与学习率等,并探索了这些参数在不同配置下对模型精度的影响。

### 2.2.1 TCN 层数的选择

从理论上讲,增加 TCN 的层数可以增强非线性表达能力,提取更深层次的特征,从而提升模型的性能。在利用本文模型进行机器阅读理解的过程中,需要对 TCN 的层数进行实验验证,验证结果如表 3 所示。从表 3 可以看出,模型的精度随着 TCN 层数的增加呈现先上升后下降的趋势,这是由于 TCN 层数越多,模型的参数量越大,反而不易训练,为了使模型能够保持良好的性能,实验选择 TCN 的层数为 3 层。

表 3 TCN 层数对模型精度的影响

Table 3 Influence of TCN layers on model accuracy

TCN 层数	精度/%
2	51.5
3	52.5
4	52.1

### 2.2.2 TCN 隐层节点数的选择

TCN 隐层节点数是模型提取数据特征能力大小的重要参数,且如果 TCN 隐层节点数太少,则网络性能较差,如果 TCN 隐层节点数太多,虽然能够减小网络的系统误差,但是会使得训练时间延长,甚至出现过拟合,因此选择合适的隐层节点数十分重要。本文考察了不同 TCN 隐层节点数对模型精度的影响,结果如表 4 所示。从表 4 可以看出,模型的精度随着 TCN 隐层节点数的增加呈现先增加后降低的趋势。因此,实验选择 TCN 隐层节点数为 450。

表 4 TCN 隐层节点数对模型精度的影响

Table 4 Influence of the number of hidden nodes in TCN on model accuracy

TCN 隐层节点数	精度/%
300	52.2
450	52.5
600	52.3

### 2.2.3 模型学习率的选择

学习率控制着参数更新的速度,如果学习率过大可能导致参数在最优解两侧振荡,如果学习率过小会极大降低模型的收敛速度。图 4 给出了本文模型在不同学习率下的精度。从图 4 可以看出,随着训练次数的增加,模型的精度逐渐趋于稳定;当学习率为 0.001 时,本文模型的精度最高,为了保证模型的收敛速度以及性能,实验将学习率设置为 0.001。

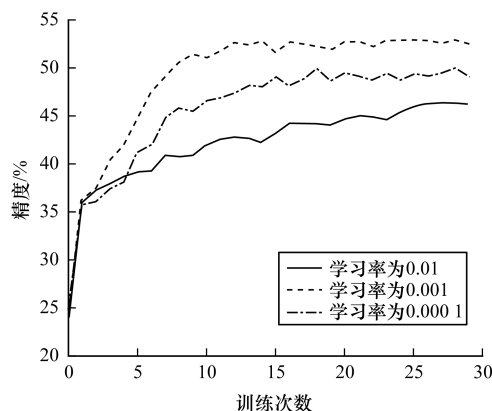


图 4 学习率对模型精度的影响

Fig. 4 Influence of learning rate on model accuracy

### 2.3 实验结果与分析

为了对模型效果进行评估,实验将 GA (Gated Attention)<sup>[8]</sup>模型、ElimiNet (Eliminate Network)<sup>[25]</sup>模型、HAF (Hierarchical Attention Flow)<sup>[16]</sup>模型、MRU (Multi-range Reasoning Units)<sup>[26]</sup>模型、HCM (Hierarchical Co-Matching)<sup>[17]</sup>模型与本文模型进行比较,并以精度作为衡量模型性能的指标,结果如表 5 所示。从表 5 可以看出,本文模型在 RACE-H 数据集与 RACE 数据集上的精度分别达到了 48.9%、52.5%,均优于其他基线模型,仅在 RACE-M 数据集上的精度次于 MRU 模型,说明了本文模型有效且具有一定的优势。

表 5 本文模型与基线模型的精度对比

Table 5 Accuracy comparison between the proposed model and the baseline model %

模型	RACE-M 数据集	RACE-H 数据集	RACE 数据集
GA 模型	43.7	44.2	44.1
ElimiNet 模型	—	—	44.7
HAF 模型	45.3	47.9	47.2
MRU 模型	57.7	47.4	50.4
HCM 模型	55.8	48.2	50.4
本文模型	56.2	48.9	52.5

为了证明本文模型中各个功能模块的有效性,本文还进行了模型简化实验来探索注意力机制、融合函数以及 TCN 对模型效果的影响,结果如表 6 所示。从表 6 可以看出,当从本文模型中去除注意力机制、融合函数或者 TCN 时,该模型的精度分别下降了 9.1%、7.8%、7.9%,下降的幅度较大,证明了注意力机制、融合函数以及 TCN 是影响模型效果的重要因素。

表 6 在 RACE 数据集上的模型简化实验结果

Table 6 Experimental results of model simplification on RACE dataset %

模型	精度
去除注意力机制模型	43.4
去除融合函数模型	44.7
去除 TCN 模型	44.6
本文模型	52.5

### 3 结束语

本文提出一种基于 TCN 的机器阅读理解 M-TCN 模型,并将其应用于多项选择任务。该模型使用注意力机制对文章、问题和候选答案进行匹配,采用 TCN 提取高层特征。实验结果表明,M-TCN 模型在 RACE 数据集上的预测精度达到了 52.5%,与同类模型相比具有一定的优势。后续将在本文工作的基础上,采用预训练技术进一步提升模型效果,并不断增强其推理能力。

#### 参考文献

- [1] ZHANG Haoyu, ZHANG Pengfei, LI Zhenzhen, et al. Self-attention based machine reading comprehension [J]. Journal of Chinese Information Processing, 2018, 32 (12): 125-131. (in Chinese)  
张浩宇, 张鹏飞, 李真真, 等. 基于自注意力机制的阅读理解模型 [J]. 中文信息学报, 2018, 32 (12): 125-131.
- [2] KARL M H, TOMAS K, EDWARD G, et al. Teaching machines to read and comprehend [C] // Proceedings of Annual Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015: 1693-1701.
- [3] XU Lili, LI Ru, LI Yuexiang, et al. Answer selection method of sentence filling for machine reading comprehension [J]. Computer Engineering, 2018, 44 (7): 183-187, 192. (in Chinese)  
徐丽丽, 李茹, 李月香, 等. 面向机器阅读理解的语句填补答案选择方法 [J]. 计算机工程, 2018, 44 (7): 183-187, 192.
- [4] LÜ Guoying, GUAN Yong, LI Ru, et al. Solution of text title selection based on neural network [J]. Computer Engineering, 2018, 44 (9): 171-176. (in Chinese)  
吕国英, 关勇, 李茹, 等. 基于神经网络的篇章标题选择求解 [J]. 计算机工程, 2018, 44 (9): 171-176.
- [5] YIN Yichun, ZHANG Ming. A neural machine reading comprehension model based on relabeling and rich features [J]. Journal of Chinese Information Processing, 2018, 32 (11): 112-116. (in Chinese)  
尹伊淳, 张铭. 一种基于数据重构和富特征的神经网络机器阅读理解模型 [J]. 中文信息学报, 2018, 32 (11): 112-116.
- [6] FELIX H, ANTOINE B, SUMIT C, et al. The goldilocks

- principle:reading children's books with explicit memory representations[C]//Proceedings of International Conference on Learning Representations. San Juan, Puerto Rico:[s. n.], 2016:228-236.
- [7] 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. Berlin, Germany: Springer, 2016:1-11.
- [8] BHUWAN D, LIU H X, YANG Z L, et al. Gated attention readers for text comprehension[C]//Proceedings of Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: Association for Computational Linguistics, 2017:1832-1846.
- [9] SORDONI A, BACHMAN P, BENGIO Y. Iterative alternating neural attention for machine reading [EB/OL]. [2019-06-20]. [https://www.researchgate.net/profile/Philip\\_Bachman/publication/303840417\\_Iterative\\_Alternating\\_Neural\\_Attention\\_for\\_Machine\\_Reading/links/59921e9daca27289539bb860/Iterative-Alternating-Neural-Attention-for-Machine-Reading.pdf](https://www.researchgate.net/profile/Philip_Bachman/publication/303840417_Iterative_Alternating_Neural_Attention_for_Machine_Reading/links/59921e9daca27289539bb860/Iterative-Alternating-Neural-Attention-for-Machine-Reading.pdf).
- [10] CHEN D Q, 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. Berlin, Germany: Springer, 2016:2359-2367.
- [11] 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. New York, USA: Association for Computational Linguistics, 2016:2383-2392.
- [12] TRISCHLER A, WANG T, YUAN X D, et al. NewsQA: a machine comprehension dataset[C]//Proceedings of the 2nd Workshop on Representation Learning for NLP. New York, USA: Association for Computational Linguistics, 2017: 191-200.
- [13] MIN J S, ANIRUDDHA K, ALI F, et al. Bidirectional attention flow for machine comprehension [EB/OL]. [2019-06-20]. <https://arxiv.org/pdf/1611.01603.pdf>.
- [14] Natural Language Computing. R-NET: machine reading comprehension with self-matching networks [C]//Proceedings of Annual Meeting of the Association for Computational Linguistics. New York, USA: Association for Computational Linguistics, 2017:1-11.
- [15] YU A W, DOHAN D, LUONG M T, et al. QANet: combining local convolution with global self-attention for reading comprehension [EB/OL]. [2019-06-20]. <https://arxiv.org/abs/1804.09541>.
- [16] ZHU Haichao, WEI Furu, QIN Bing, et al. Hierarchical attention flow for multiple-choice reading comprehension [C]//Proceedings of AAAI Conference on Artificial Intelligence. New Orleans, USA:[s. n.], 2018:6077-6085.
- [17] WANG Shuohang, YU Mo, CHANG Shiyu, et al. A co-matching model for multi-choice reading comprehension [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. New York, USA: Association for Computational Linguistics, 2018:746-751.
- [18] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. [2019-06-20]. <https://arxiv.org/abs/1803.01271>.
- [19] LAI Guokun, XIE Qizhe, LIU Hanxiao, et al. RACE: large-scale ReAding comprehension dataset from examinations[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen. New York, USA: Association for Computational Linguistics, 2017:785-794.
- [20] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [21] CHEN Q, ZHU X, LING Z H, et al. Natural language inference with external knowledge[EB/OL]. [2019-06-20]. <https://arxiv.org/pdf/1711.04289.pdf>.
- [22] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016:770-778.
- [23] SALIMANS T, KINGMA D P. Weight normalization: a simple reparameterization to accelerate training of deep neural networks[C]//Proceedings of Annual Conference on Neural Information Processing Systems. Barcelona, Spain: MIT Press, 2016:901-906.
- [24] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
- [25] PARIKH S, SAI A, NEMA P, et al. ElimiNet: a model for eliminating options for reading comprehension with multiple choice questions [C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California, USA: International Joint Conferences on Artificial Intelligence Organization, 2018:4272-4278.
- [26] TAY Y, TUAN L, HUI S. Multi-range reasoning for machine comprehension [EB/OL]. [2019-06-20]. <https://arxiv.org/abs/1803.09074>.