

# Ti-Reader: 基于注意力机制的藏文机器阅读理解端到端网络模型

孙媛<sup>1,2,\*</sup> 陈超凡<sup>1,2</sup> 刘思思<sup>1,2</sup> 赵小兵<sup>1,2</sup>

<sup>1</sup>中央民族大学 信息工程学院, 北京 100081

<sup>2</sup>国家语言资源监测与研究少数民族语言中心

\*Corresponding author: Yuan Sun

tracy.yuan.sun@gmail.com, chaofanalex@qq.com,

liusisi.s@qq.com, nmzxb\_cn@163.com

## 摘要

机器阅读理解旨在教会机器去理解一篇文章并且回答与之相关的问题。为了解决低资源语言上机器阅读理解模型性能低的问题, 本文提出了一种基于注意力机制的藏文机器阅读理解端到端网络模型Ti-Reader。首先, 为了编码更细粒度的藏文文本信息, 本文将音节和词相结合进行词表示, 然后采用词级注意力机制去关注文本中的关键词, 采用重读机制去捕捉文章和问题之间的语义信息, 采用自注意力机制去匹配问题与答案的隐变量本身, 为答案预测提供更多的线索。最后, 实验结果表明, Ti-Reader模型提升了藏文机器阅读理解的性能, 并且在英文数据集SQuAD上也有较好的表现。

**关键词:** 机器阅读理解; 注意力机制; 端到端网络; 藏文

## Ti-Reader: An End-to-End Network Model Based on Attention Mechanisms for Tibetan Machine Reading Comprehension

Yuan Sun<sup>1,2,\*</sup> Chaofan Chen<sup>1,2</sup> Sisi Liu<sup>1,2</sup> Xiaobing Zhao<sup>1,2</sup>

<sup>1</sup> Minzu University of China, Beijing 100081

<sup>2</sup> National Language Resource Monitoring & Research Center of Minority Languages

\*Corresponding author: Yuan Sun

tracy.yuan.sun@gmail.com, chaofanalex@qq.com,

liusisi.s@qq.com, nmzxb\_cn@163.com

## Abstract

Machine reading comprehension aims to teach machines to understand an article and answer related questions. To solve the problem of low performance of machine reading comprehension models in low-resource languages, this paper proposes an end-to-end network model based on attention mechanisms named Ti-Reader for Tibetan machine reading comprehension. First, to encode more fine-grained Tibetan text information, this paper combines syllables and words for word embedding, and then uses word-level attention to pay more attention to the keywords in the article. Moreover, the re-read mechanism is used to capture the semantic information between the article and the questions, and the self-attention is used to match the hidden variables of the question and the answer, which provides more clues for giving answer. Finally, the experimental results show that Ti-Reader improves the performance of Tibetan machine reading comprehension, and it has also a good performance on the English dataset SQuAD.

**Keywords:** Machine reading comprehension, Attention, End-to-end network, Tibetan

## 1 引言

机器阅读理解是自然语言处理领域从感知文本到理解文本的重要一步，它要求机器阅读文本信息如文章或故事，然后回答与该文本相关的问题。这些问题可以被设计成各式各样的形式用于测试机器的理解能力。在早期，大多数机器阅读理解系统是基于规则或统计模型，研究人员必须手工设计一些复杂的语法或语义规则。这些系统只能达到30%-40%的精度(Riloff and Thelen, 2000)，因此这些成果并没有引起广泛的关注。在接下来的几十年中，研究人员开始关注机器阅读理解数据集的构建，Hermann等人提出了一个完形填空式的英语机器阅读理解数据集-CNN&Daily Mail。Hill等人发布了Children's Book Test 数据集(Hill et al., 2015)。Lai等人(Lai et al., 2017)在2017年发布了RACE数据集。Rajpurkar等人于2016年(Rajpurkar et al., 2016)公布了Stanford Question answer dataset (SQuAD)的大规模问答数据集，它是抽取式的机器阅读理解数据集，数据集的问题是人工提出的，答案是文章中出现的连续的一段片段，包含了107,785个高质量的问题。随着这些数据集的发布，一些优秀的英文机器阅读理解模型如S-Net(Tan et al., 2018)、AS Reader(Tan et al., 2016)、IA Reader(Sordoni et al., 2016)等被提出。这些模型在英文阅读理解任务上取得了和人类相媲美的精确匹配性能。

然而，对于藏文等低资源语言的机器阅读理解却鲜有提及。其一是因为与英语机器阅读理解相比，现有的藏文分词工具可能会带来错误，产生语义歧义，从而将错误传播到下游任务中。其二是对于低资源的机器阅读理解任务，在小规模的数据集上很难获得良好的性能，因此需要机器阅读理解模型来加强模型的理解力。针对以上问题，本文提出了一种端到端的藏文机器阅读理解模型。为了减少分词引起的错误传播，模型引入了音节级别的信息。另外，为了提高模型的理解能力，本文采用了一种多级注意力机制。

综上所述，本文的主要贡献如下：

(1) 为了编码更细粒度的藏文文本信息，本文提出了一种将音节和词相结合的词表示方法，使模型能够学习藏文中更复杂的信息，降低分词带来的错误语义信息。

(2) 藏文的文本较长，段落中会存在一些与问题无关的干扰信息，这些噪音信息会影响模型预测的准确率，因此本文采用基于词级别的注意力机制去关注与答案相关的关键词，从而避免长文本信息中无关信息的干扰。

(3) 为了加强模型的理解能力，本文采用重读机制去捕捉文章和问题之间的语义信息，采用自注意力机制去匹配问题与答案的隐变量本身，为答案预测提供更多的线索。

## 2 相关研究

近年来，随着许多英文大规模数据集的出现，基于深度学习的机器阅读理解研究取得了一些显著的成绩。Wang等人提出Match-LSTM模型(Wang and Jiang, 2016)，他们分别采用长短时记忆网络(LSTM) (Hochreiter and Schmidhuber, 1997)对问题和段落进行编码，然后在LSTM单元中引入基于注意力的问题加权表示。随后，微软团队为了捕捉文章中单词之间的长期依赖关系提出了R-Net模型(Wang et al., 2017)，通过引入额外的自注意力层来实现，实验结果表明通过引入自注意力机制能够提高模型的准确性。Cui等人提出了“Attention over Attention”阅读器模型(Cui et al., 2016)，这是一种基于行和列的相结合的注意计算方法。为了进一步提高模型的准确性，他们采用了“N-Best”和“重新排列”的策略来验证答案。与以往的工作不同，Seo等人采用了两个方向的注意力并提出了BiDAF模型(Seo et al., 2016)，他们认为通过计算问题到文章和文章到问题两个方向的注意力可以提取更多的信息，BiDAF没有将文本信息嵌入到固定长度的向量，而是使信息在向量之间流动以减少早期信息加权求和而导致的信息损失。另外，他们不直接依赖上一时刻的注意力，这使得后续的注意力计算不受先前错误的注意力信息的影响，最后他们的实验也证明了该模型的有效性。Xiong等人则提出了一个DCN模型(Xiong et al., 2016)，该模型使用交互注意力机制来捕捉问题和段落之间的交互。DCN的特点在于应答层，它们使用多轮迭代指向机制，每一轮迭代都会产生答案开始和结束位置的预测。基于这两个预测，采用LSTM和Highway Max-out网络用于更新新一轮的答案范围预测。在交互层，DCN使用双向注意力机制来计算原始的查询感知表示。

上述模型由于注意力层数较少以及网络深度不够深的结构特点，因此存在捕获问题和段落之间语义信息较弱的问题。为了解决这个问题，最近一系列的工作通过堆叠多个注意力层来增

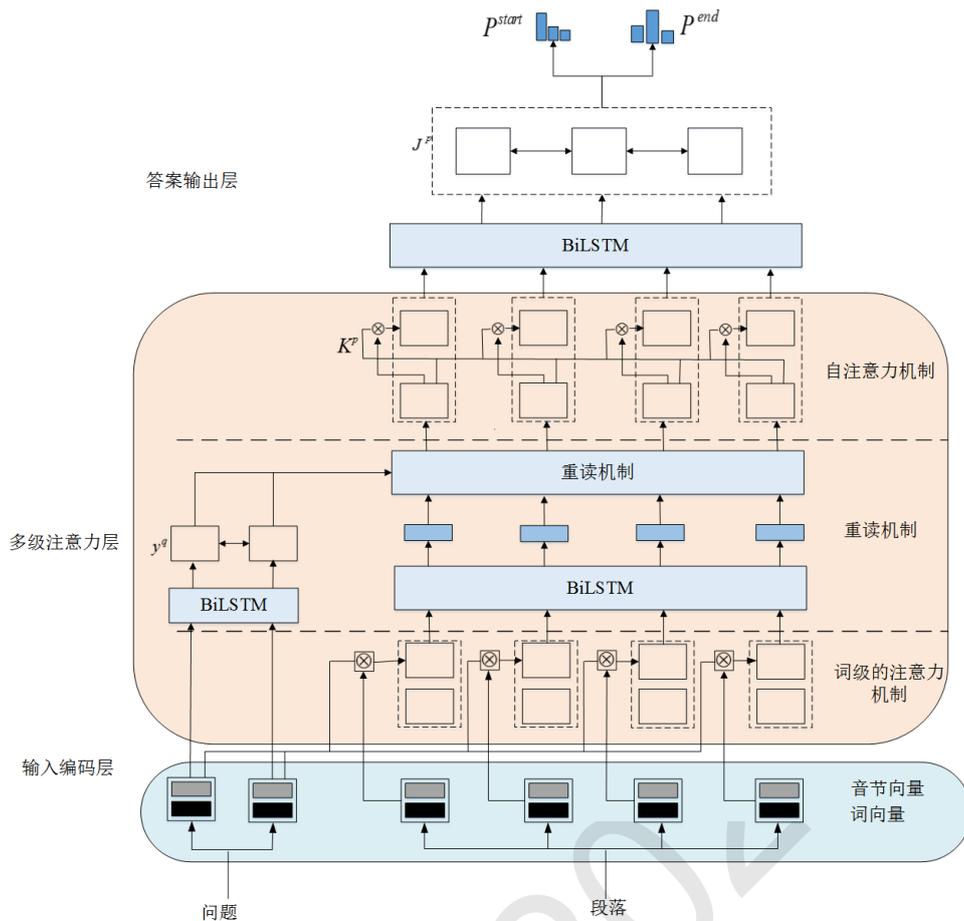


图 1: Ti-Reader模型架构

强模型的理解能力。Huang等(Huang et al., 2017)提出Fusion-Net, 该模型使用全连接感知的多层注意力结构来获取问题中完整的语义信息并将其集成到段落表示中。Wang等人(Wang et al., 2018)则是提出了一种多粒度的分层注意力融合网络。通过计算不同粒度下的注意力分布然后进行分层语义融合, 他们的实验证明了多层注意力能够获取问题与文章之间的交互信息, 从而提高模型的性能。Tan等人(Tan et al., 2018)提出提取生成模型, 他们使用RNN和注意力机制来构建问题和上下文表示, 然后使用seq2seq模型生成基于关键信息的答案。

同时, 为了获得问题和段落的语义表示, 需要对嵌入表示进行编码, 以捕获上下文语义信息。常见的编码组件包括递归神经网络 (RNN) 和卷积神经网络 (CNN)。RNN可以捕获句子之间的远距离依存关系, 大多数研究广泛使用RNN作为嵌入网络来编码段落和问题(Weston et al., 2014)。这些工作可以有效地编码语义信息, 但是需要很长时间来训练它们的模型。为了解决这个问题, 一些研究人员提议将CNN应用于机器阅读理解(Yin et al., 2016)。CNN可以有效地提取本地信息, 与RNN相比, CNN的优势在于可以并行训练, 因此速度更快。但是CNN的主要缺点是它只能捕获本地信息特征, 而不能有效地处理长距离依赖性, 因此CNN必须解决长距离信息提取的问题。本文在以上工作的基础上, 提出了一种基于注意力机制的藏文机器阅读理解端到端网络模型Ti-Reader, 旨在解决低资源语言(如藏文)机器阅读理解的问题。

### 3 模型总体架构

模型的总体框架如图1所示, 主要包括: 融入音节信息的输入嵌入层、多级注意力层、答案输出层。首先, 输入嵌入层针对输入的藏文段落和问题文字信息进行编码, 并融入细粒度的音节信息。然后, 为了解决模型理解能力不足的问题, 本文将文本词向量送入多级注意机制层, 这部分包括词级注意力机制、重读机制和自注意机制, 分别用于对段落中与问题相关的关键词搜索、对段落和问题之间的关键语义信息的捕捉和对遗漏的关键线索进行搜索。最后答案输出

层输出预测答案的始末位置。

### 3.1 数据预处理

与英文不同，藏文是拼音文字，字级别的单位是音节。一些音节可以指示一些有意义的“格”。藏文中的“格”是一种功能音节，可以区分字并解释字在短语或句子中的作用。它可以帮助模型给出正确的答案。实际上，藏文中有许多音节可以像“格”一样为机器阅读理解任务提供一些关键信息。因此，有必要将音节信息嵌入编码层中。另一方面，音节的嵌入可以减少由于不正确的分词而造成的语义歧义。基于以上考虑，本文将音节和单词信息进行了结合。接下来，我们将介绍音节级和词级的藏文文本预处理。

#### (1) 音节级别的预处理

因为音节之间存在定界符，所以分割音节很容易，可以根据定界符“.”分隔音节。

#### (2) 词级别的预处理

对于词级别切分，本文使用藏文词切分工具(龙从军 et al., 2015)。基于词级以及音节级的文本预处理的具体格式如表1所示。

原始句子	ལུ་སྐྱེད་ཟས་སྣོད་མང་ཆེ་བ་ནི་བསྐྱར་སྐྱེས་ཀྱི་ཚྲོས་འབྱེད།
音节切分	ལུ / སྐྱེད / ཟས / སྣོད / མང / ཆེ / བ / ནི / བསྐྱར / སྐྱེས / ཀྱི / ཚྲོས / འབྱེད
词级切分	ལུ་སྐྱེད་ / ཟས་སྣོད་ / མང་ཆེ་བ་ / ནི་ / བསྐྱར་སྐྱེས་ / ཀྱི་ / ཚྲོས་ / འབྱེད།

表 1: 数据预处理样例

### 3.2 输入编码层

藏文具有很强的语法规则，它是由一个一个的音节组成，因此音节是藏文的基本构成单位。值得注意的是，某些音节可能包含丰富的语义信息，这些信息将有助于预测正确答案。因此，在输入编码层我们将音节嵌入词向量中，从而可以为网络提取更多的语义信息。

假设有一个问题序列 $Q=\{q_1, q_2, q_3, \dots, q_n\}$ 和一个段落序列 $P=\{p_1, p_2, p_3 \dots, p_m\}$ ，我们将它们分别转换为音节级和词级的嵌入( $\{s_1^q, s_2^q, s_3^q, \dots, s_n^q\}$  和  $\{s_1^p, s_2^p, s_3^p, \dots, s_m^p\}$ )。同时，我们使用预训练模型对问题和段落进行编码。每个单词通过fastext编码为100维向量。对于音节级编码，我们使用双向长短期记忆神经网络 (BiLSTM)，并将最终状态向量作为音节级的表示。最后，我们通过两层高速公路网络融合两个不同级别的向量 (Srivastava et al., 2015)，文章和问题最终编码形式化表示为 $\{M_t^q\}_{t=1}^n$  和  $\{M_t^p\}_{t=1}^m$ 。

### 3.3 多级注意力机制

#### 3.3.1 词级别的注意力机制

就像人们做阅读理解题目一样，人们首先将阅读问题，然后开始快速地阅读文章，标记出与问题相关的单词并重点去关注这些关键词，最后搜索正确的答案。受此启发，本文提出了一种词级注意力进行单词级别的关注，并计算段落中每个单词与问题的重要性。假设通过嵌入层的词语的嵌入表示为 $\{M_t^p\}_{t=1}^m$ ，而问题词级嵌入为 $\{M_t^q\}_{t=1}^n$ 。文章中每个词的注意力权重可以由公式 (1) 计算得到。

$$S_u = V^T * \tan(W_u^Q * M_i^q + W_u^P * M_j^p) \quad (1)$$

其中， $W_u^q$  和  $W_u^p$  是可以训练的权重矩阵， $S_u$  是相关性矩阵。接下来我们对这个相关性矩阵 $S_u$ 中每一行进行softmax函数归一化，其形式化表示为公式 (2) 所示。

$$a_u \propto \exp(S_u) \quad (2)$$

为了确定段落中哪些单词有助于回答该问题，我们计算具有问题意识的分配矩阵，其计算公式如 (3) 所示。

$$A_i^p = \sum a_u M_j^q \quad (3)$$

最后，我们使用双向的长短期记忆网络去获取句子级别的文章表示，其形式化表示为公式(4)所示。

$$V_t^p = BiLSTM(V_{t-1}^p, [A_t^p, M_t^p]) \quad (4)$$

### 3.3.2 重读机制

词级别的注意力层是浅层注意力计算。为了提高模型性能，本文采用高层级别的注意力机制来考虑哪个句子包含正确的答案。因此，本文引入了“重读注意力机制”。重读注意力机制旨在计算句子级别上段落和问题之间的注意力。在计算注意力之前，我们需要理解问题。即对于问题中的每个词，本文使用双向的长短期记忆网络去生成高层的问题表示 $y_t^q$ ，其形式化表示如公式(5)所示。

$$y_t^q = BiLSTM(y_{t-1}^q, [s_t^q, w_t^q]) \quad (5)$$

这里的 $y_{t-1}^q$ 表示的是前一个状态的隐藏向量， $s_t^q$ 是输入嵌入层中问题的音节嵌入的输出， $w_t^q$ 则是词级别注意力机制层的输出。基于对问题的理解，同样的，本文使用重读注意力机制，其计算公式如公式(6)-(8)所示。

$$S_v = V^T \tan(W_v^Q y_t^q + W_v^P V_j^p) \quad (6)$$

$$a_v \propto \exp(S_v) \quad (7)$$

$$A_i^p = \sum a_v y_i^q \quad (8)$$

其中， $S_v$ 是文章和问题之间语义相关性矩阵， $y_i^q$ 是问题编码的语义向量， $V_j^p$ 是词级别注意力机制层的输出向量。

最后，本文使用双向的长短期记忆网络对重读机制层的输出进行编码，输出向量表示为 $K_t^p$ ，其计算公式如公式(9)所示。

$$K_t^p = BiLSTM(K_{t-1}^p, [A_t^p, y_t^q]) \quad (9)$$

### 3.3.3 自注意力机制

通过单词级别的注意力机制和语义级别的重读注意力机制，可以根据单词和问题中的语义信息找到答案。但是，答案和问题的语义和结构形式之间不可避免地存在一定差异，这可能导致在段落和问题之间的交互过程中丢失了一些重要的信息，从而导致答案的预测不准确。为了解决这个问题，本文引入了一种自注意力机制，即匹配问题与段落的隐变量本身，以便可以动态调整答案。整个过程可以描述为公式(10)-(12)。

$$S_j = V^T \tanh(W_h^p K_i^p + W_h^p K_j^p) \quad (10)$$

$$a_j \propto \exp(S_j) \quad (11)$$

$$A_i^p = \sum a_j K_i^q \quad (12)$$

这里 $S_j$ 是相关性矩阵， $A_i^p$ 是注意力权重的隐向量。接下来自注意力机制向量被送入另外一个双向长短期注意力机制网络去生成最后的向量表示 $\{J_t^p\}_{t=1}^m$ ，其形式化表示如公式(13)所示。

$$J_t^p = BiLSTM(J_{t-1}^p, [A_t^p, K_t^p]) \quad (13)$$

其中， $J_{t-1}^p$ 是网络中前一个状态的隐层向量， $K_t^p$ 是重读注意力机制层的输出。

### 3.4 答案输出层

该层的主要目标是预测答案的起始位置。本文使用softmax层来实现，该层可以预测给定段落中每个位置作为答案的开始或结束的概率，如公式(14)和(15)所示。

$$p^{start} = \text{soft max}(W_1 J^p) \quad (14)$$

$$p^{end} = \text{soft max}(W_2 J^p) \quad (15)$$

其中， $W_1$ 和 $W_2$ 是可以被训练的参数矩阵， $p^{start}$ ， $p^{end}$ 是答案的起始位置。

## 4 实验结果及分析

### 4.1 数据集

本文在SQuAD1.0和TibetanQA上进行了实验。表2是这些数据集的统计信息。

数据集	训练集		测试集	
	段落	问题	段落	问题
SQuAD	17,007	68,758	1,889	18,841
TibetanQA	5,194	16,000	587	4,000

表 2: 数据集的统计结果

SQuAD1.0(Rajpurkar et al., 2016): 这是斯坦福学校在2016年通过“众包”方式人工构建的一种抽取式阅读理解数据集, SQuAD1.0拥有超过100,000个问答对。数据集的文章段落摘自于维基百科中536篇实体知识文章, 每个问题的答案都是一段连续的文章片段。

TibetanQA(孙媛 et al., 2021): 这是我们前期构建的藏文机器阅读理解数据集, 它同样是采用“众包”方式人工构建的数据集。该数据集的文章选自云藏网, 总计收集了1,513篇各领域知识实体的文本, 并构建了20,000个问题答案对。

### 4.2 评估指标

为了评估模型的效果, 本文使用两种评估指标“完全匹配 (EM)”和“F1值”。

EM: 计算模型的真实答案和预测答案之间的匹配程度。对于机器阅读理解预测任务, 如果答案的位置与标签匹配, 则正确的数字将增加1, 否则不进行计数。例如, 有 $m$ 个问题, 如果模型可以正确回答 $n$ 个问题, 则采用EM可以按照公式 (16) 进行计算。

$$EM = \frac{n}{m} \quad (16)$$

F1值: 我们将预测答案和标准答案视为字符数据集, 计算字符集级别的精度和召回率的平均值。字符级别的真正例 (TP) 表示预测答案和标准答案之间的字符相同, 而字符级别的假正例 (FP) 是不在标准答案中但在预测答案中的字符。假反例 (FN) 是存在标准答案中的字符, 而不存在预测答案中的字符。 $N(TP)$ 表示字符级别真正例的数量,  $N(FP)$ 表示字符级别的假正例的数量,  $N(FN)$ 则是字符级别的假反例的数量。最后, 我们对所有问题取平均值以计算模型的F1值, 如公式 (17) - (19) 所示。

$$precision = \frac{N(TP)}{N(TP) + N(FP)} \quad (17)$$

$$recall = \frac{N(TP)}{N(TP) + N(FN)} \quad (18)$$

$$F1 = 2 \frac{precision \cdot recall}{precision + recall} \quad (19)$$

### 4.3 在不同模型上的实验结果分析

本文采用R-Net、BiDAF和QANet作为基线模型, 这些都是十分经典的模型, 它们在英文机器阅读理解数据集上都有着出色的表现。

(1) R-Net: 此模型由微软亚研究院团队提出(Wang et al., 2017), 他们通过问题注意力机制和门禁注意力网络提升模型的效果。

(2) BiDAF: BiDAF模型由Seo等人提出(Seo et al., 2016), 与R-Net不同, BiDAF模型采用两个方向的交互层。他们不像R-Net那样使用自匹配, 而是计算关于查询到上下文和上下文到查询的两个注意事项。

(3) QANet: 该模型由Adam等人提出(Yu et al., 2018), 他们将局部卷积与全局自注意力结合在一起, 并在SQuAD数据集上获得了更好的性能。

No	Model	SQuAD		Our dataset	
		EM	F1	EM	F1
1	R-Net	71.3	79.7	55.8	63.4
2	BiDAF	68.0	77.3	58.6	67.8
3	QANet	<b>73.6</b>	<b>82.7</b>	57.1	66.9
4	Ti-Reader	73.1	81.2	60.2	70.7
5	Ti-Reader + syllable	-	-	<b>67.9</b>	<b>77.4</b>

表 3: 不同模型在两种数据集上的实验结果

No	Model	EM	增加	F1	增加
1	R-Net	55.8	-	63.4	-
2	R-Net+syllable	60.9	+5.1	68.8	+5.4
3	BiDAF	58.6	-	67.8	-
4	BiDAF+syllable	62.2	+3.6	71.9	+4.1
5	QANet	57.1	-	66.9	-
6	QANet+syllable	61.9	+4.8	72.1	+5.2
7	Ti-Reader	60.2	-	70.7	-
8	Ti-Reader+syllable	<b>67.9</b>	<b>+7.7</b>	<b>77.4</b>	<b>+6.7</b>

表 4: 在TibetanQA数据集上模型的输入嵌入消融实验

考虑到英文中没有音节，本文在SQuAD上的模型删除了音节嵌入。接下来，本文对SQuAD和TibetanQA数据集进行一些实验。所有模型均使用fastText嵌入并由本文实现，实验结果如表3所示。

可以发现Ti-Reader在两个不同数据集上都有较好的性能。对于SQuAD，Ti-Reader的EM和F1值分别达到73.1%和81.2%。与R-Net相比，Ti-Reader在EM上增加了1.8%，在F1值上增加了1.5%，与BiDAF相比，Ti-Reader在F1值上增加了3.9%。对于TibetanQA数据集，可以发现Ti-Reader优于其他模型，EM值达到60.2%，F1值达到70.7%，与R-Net相比，Ti-Reader在EM上增加了4.4%，在F1值上增加了7.3%，与BiDAF相比，Ti-Reader在F1值上增加了2.9%，与QANet相比，Ti-Reader在F1值上增加了3.8%。当我们使用音节级别的嵌入时，Ti-Reader模型在EM上提升了7.7%，在F1值上提升了6.7%。

#### 4.3.1 输入嵌入层对实验结果的影响

在上一节中，我们说明了藏文中的许多音节可以提供一些语义信息，这有助于机器更好地理解句子信息。本文加入音节信息后进行实验，结果如表4所示。

对于R-Net模型，BiDAF模型和QANet模型，融入音节信息后EM值和F1值均有所提升，其中Ti-Reader模型的提升最大，在EM上提升了7.7%，在F1值上提升了6.7%，这表明音节嵌入可以提高藏文机器阅读理解任务模型的性能。主要的原因是：首先，缺乏音节级别的嵌入会缺少一些语义信息。其次，词级别的信息虽然可以更好地表示语义信息，但是不能解决未登录词的问题。在实验中，我们确实发现某些单词不在词汇表中，这些单词被标记为“UNK”。这些问题都会降低模型的性能。

#### 4.3.2 多级注意力机制对实验结果的影响

此外，我们针对多级注意力机制（词级别的注意力机制、重读机制和自注意机制）进行了实验，结果如表5所示。

当去掉词级别的注意力机制（-WA）时，模型的效果有所下降，其中EM值降低了7.4%，F1值降低了5.6%。这说明词级别的注意力机制可以动态分配每个词的权重，能够使模型专注于那些有价值的词，从而提高模型的性能。

自注意力机制首先要阅读段落中的句子，然后对上下文信息进行权重的分配，并提供预测答案的线索。通过再次的阅读文章，模型可以找到丢失的重要信息。当去掉自注意力机制（-SA）时，模型的EM下降了4.7%，F1值下降了4.3%。

重读机制实现了段落和问题之间的相关性计算。可以发现，当去掉重读机制（-RA）时，模型的EM下降了9.1%，F1值下降了8.2%。为了更好地说明重读机制的作用，本文绘制了采用重读机制之前和之后的文章词的权重分布图，如图2，图3所示。

No	Model	EM	Increase	F1	Increase
1	Ti-Reader	67.9		77.4	
2	-WA	60.5	-7.4	71.8	-5.6
3	-SA	63.2	-4.7	73.1	-4.3
4	-RA	58.8	-9.1	69.2	-8.2

表 5: 模型的消融实验

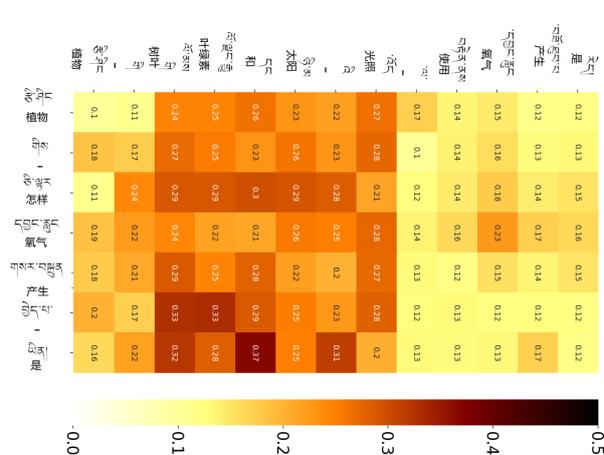


图 2: 重读机制之前的权重分布

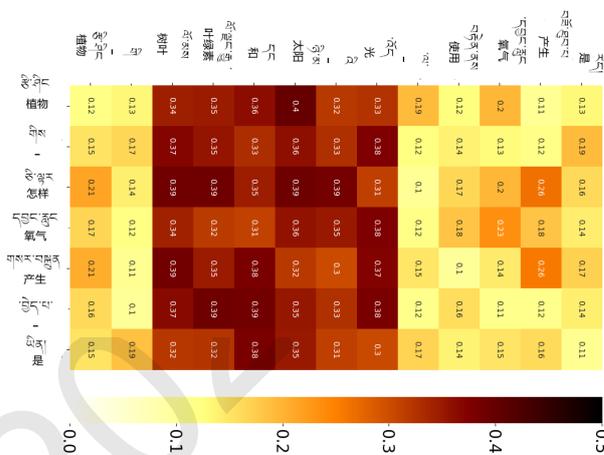


图 3: 重读机制之后的权重分布

从图2和图3可以看出，针对句子“ $\text{ཆེ་གི་ལོ་མཁའ་ལོ་རྩལ་རྒྱ་དང་ཉེ་མའི་འོད་ལ་བརྟེན་ནས་དབྱེད་སྤྲུང་བཟོ་སྐབས་པ་རེད།}$ ”(植物通过叶绿素和阳光制造氧气.....)，当被问到“ $\text{ཆེ་གི་ལོ་མཁའ་ལོ་རྩལ་རྒྱ་དང་ཉེ་མའི་འོད་ལ་བརྟེན་ནས་ཅི་ལྟར་དབྱེད་སྤྲུང་བཟོ་སྐབས་པ་རེད།}$ ”(植物是如何制造氧气的?)时，容易发现正确的答案“ $\text{འོ་མཁའ་ལོ་རྩལ་རྒྱ་དང་ཉེ་མའི་འོད་}$ ”(叶绿素和阳光)，其权重在重读机制前后有着明显增加。实验结果表明，重读机制可以捕获段落中的关键信息。

#### 4.4 不同长度的文章编码对结果的影响

此外，本文进一步探究了句子长度对模型效果的影响，其实验数据如表6所示。

文章长度 (词)	EM	F1
100	55.4	61.2
200	58.7	64.5
300	60.2	67.5
400	61.4	68.2
500	<b>67.9</b>	<b>77.4</b>
600	48.7	56.4
700	57.2	65.7

表 6: 不同长度的文章编码实验结果

实验结果表明，文章的长度也是影响模型效果的一个重要因素。从表6中可以看出，当段落落在300到500个单词时，F1值稳定在60%左右。而当文章长度在500个词时，该模型表现出最好的结果。另外，当文章的长度超过500时，F1值将受到影响呈现下降的趋势。

#### 4.5 不同类型的答案预测结果分析

答案类型	EM	F1
数字	<b>74.5</b>	<b>82.1</b>
人物	47.2	54.1
地理位置	63.7	72.8
组织	55.5	61.4
短语	52.5	61.5
日期	70.2	81.3

表 7: 不同的答案类型模型预测的准确率

我们分析了不同答案类型模型预测的准确率，如表7所示。针对数字类型的答案，我们的模型取得的准确率最高，其中F1达到82.1%。对于日期，我们模型的F1值可以达到81.3%。短语类型可以达到61.5%。但是针对人物类型的答案预测却不够理想。我们认为影响人物答案类型的预测主要原因是：（1）在分词阶段，藏文人名没有被正确切分。（2）藏文名字的构成比较复杂，机器理解相对困难。

## 5 结论与展望

本文提出一种基于注意力机制的藏文机器阅读理解端到端网络模型Ti-Reader。该模型使用了分层注意力机制，包括词级别的注意力机制，重读机制和自注意力机制。同时，进行了一些消融实验，并证明了它们的有效性。另一方面，我们通过对比三种经典的英文机器阅读理解模型BiDAF, R-Net和QANet，实验表明Ti-Reader对藏文机器阅读理解任务具有更好的表现。但是我们的模型针对一些问句仍然不能给出正确的答案。未来的工作，我们一方面将探索模型的可解释性，并对模型的常识推理提出更高要求，另一方面进一步提高模型的准确性。

## 致谢

本论文得到了国家自然科学基金项目（61972436）资助。

## 参考文献

- Yiming Cui, Zhipeng Chen, Si Wei, Shijin Wang, Ting Liu, and Guoping Hu. 2016. Attention-over-attention neural networks for reading comprehension. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.
- Felix Hill, Antoine Bordes, Sumit Chopra, and Jason Weston. 2015. The goldilocks principle: Reading children’s books with explicit memory representations. *Conference and Workshop on Neural Information Processing Systems*.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Hsin-Yuan Huang, Chenguang Zhu, Yelong Shen, and Weizhu Chen. 2017. Fusionnet: Fusing via fully-aware attention with application to machine comprehension. *arXiv preprint arXiv:1711.07341*.
- Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. 2017. Race: Large-scale reading comprehension dataset from examinations. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392.
- Ellen Riloff and Michael Thelen. 2000. A rule-based question answering system for reading comprehension tests. In *ANLP-NAACL 2000 Workshop: Reading Comprehension Tests as Evaluation for Computer-Based Language Understanding Systems*.

- Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. 2016. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603*.
- Alessandro Sordani, Philip Bachman, Adam Trischler, and Yoshua Bengio. 2016. Iterative alternating neural attention for machine reading. *arXiv preprint arXiv:1606.02245*.
- Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Highway networks. *arXiv preprint arXiv:1505.00387*.
- Chuanqi Tan, Furu Wei, Nan Yang, Bowen Du, Weifeng Lv, and Ming Zhou. 2016. Text understanding with the attention sum reader network. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 908–918.
- Chuanqi Tan, Furu Wei, Nan Yang, Bowen Du, Weifeng Lv, and Ming Zhou. 2018. S-net: From answer extraction to answer synthesis for machine reading comprehension. In *the Association for the Advance of Artificial Intelligence*.
- Shuohang Wang and Jing Jiang. 2016. Machine comprehension using match-lstm and answer pointer. *International Conference on Learning Representations*.
- Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou. 2017. Gated self-matching networks for reading comprehension and question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 189–198.
- Wei Wang, Ming Yan, and Chen Wu. 2018. Multi-granularity hierarchical attention fusion networks for reading comprehension and question answering.
- Jason Weston, Sumit Chopra, and Antoine Bordes. 2014. Memory networks. *arXiv preprint arXiv:1410.3916*.
- Caiming Xiong, Victor Zhong, and Richard Socher. 2016. Dynamic coattention networks for question answering. *arXiv preprint arXiv:1611.01604*.
- Wenpeng Yin, Sebastian Ebert, and Hinrich Schütze. 2016. Attention-based convolutional neural network for machine comprehension. *Proceedings of the Workshop on Human-Computer Question Answering*.
- Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V Le. 2018. Qanet: Combining local convolution with global self-attention for reading comprehension. *arXiv preprint arXiv:1804.09541*.
- 孙媛, 刘思思, 陈超凡, 旦正错, and 赵小兵. 2021. 面向机器阅读理解的高质量藏语数据集构建. 第二十届中国计算语言学大会 (*The Twentieth China National Conference on Computational Linguistics, CCL 2021*) .
- 龙从军, 刘汇丹, 诺明花, and 吴健. 2015. 基于藏语字性标注的词性预测研究. *中文信息学报*, 29(5):211–216.