

结合实体感知增强与外部知识的阅读理解

乐远, 张宇, 刘挺

(哈尔滨工业大学社会计算与信息检索研究中心, 黑龙江 哈尔滨 150001)

摘要: 人类阅读理解和机器阅读理解一个很大的差异是, 人类很善于利用除了文本之外的一些外部知识, 来辅助自己理解获取答案。本文在预训练语言模型 Bert 的基础之上, 进一步引入外部知识来提升机器阅读理解效果, 考虑到当前的预训练模型对文本进行分词后会将文本中的部分实体词拆开, 而引入的外部知识都是实体级别的, 这将导致在融合外部知识的过程中引入很大的噪声, 感知实体能力弱, 进而影响机器阅读理解融合外部知识的能力。因此, 针对这个问题, 我们提出了结合实体感知增强与外部知识的机器阅读理解模型 FSNER-net, 该模型联合命名实体识别任务提高了机器阅读理解的实体感知能力, 进一步增强了机器阅读理解融合外部知识的能力, 最后在英文阅读理解数据集 ReCoRD 的实验结果上证明了我们方法的有效性。

关键词: 机器阅读理解; 外部知识; 命名实体识别; 预训练模型

中图分类号: TP391

文献标识码: A

基金号: 2019YFF0303003, 面向冬奥场景的多语种智能问答关键技术研究, 国家科技部重点研发计划; 61976068, 面向智能客服的问题语义分析相关技术研究, 国家自然科学基金项目

Combining entity perception enhancement with external knowledge reading comprehension

Yuan Le, Yu Zhang, Ting Liu

(Research Center for Social Computing and Information Retrieval, Harbin Institute of
Technology, Harbin, Heilongjiang 150001, China)

Abstract: A big difference between human reading comprehension and machine reading comprehension is that human beings are very good at using some external knowledge other than text to assist themselves in understanding and obtaining answers. Based on the pre-trained language model Bert, this paper further introduces external knowledge to improve the effect of machine reading comprehension. Considering that the current pre-trained model splits the text, some entity words in the text will be disassembled, and the external Knowledge is at the entity level, which will lead to a lot of noise in the process of fusing external knowledge, weak ability to perceive entities, which will affect the ability of machine reading to understand external knowledge. Therefore, in response to this problem, we propose a machine reading comprehension model FSNER-net that combines entity perception enhancement and external knowledge. This model and named entity recognition task improve the entity perception ability of machine reading comprehension and further enhance the integration of machine reading comprehension with external the ability of knowledge finally proves the effectiveness of our method on the experimental results of the English reading comprehension data set ReCoRD.

Key words: Machine Reading Comprehension; External Knowledge; Named Entity Recognition; Pretrained Models

1 引言

机器阅读理解是自然语言处理乃至 AI 界前沿的一个火热话题, 它要求机器能够根据给定的文本, 来回答问题的答案, 以此来衡量机器对于自然语言的理解能力。近几年, 随着深度学习在自然语言处理中的迅猛发展, 特别是预训练语言模型 Bert^[1]等的出现, 使得很多机器阅读理解任务的效果达到了一个新的高度, 这其中很重要的原因就是预训练语言模型借助优秀的深度学习架构, 能够很好地利用大量的无标注文本, 从而能够获得文本更深层的语

义信息。

然而在现实世界的很多实际应用当中，很多阅读理解的文本内容很复杂，涉及到很多复杂的背景知识，仅仅通过给定的文本内容并不能够很好的去理解，从而很难去准确的回答问题的答案，这就导致很多机器阅读理解模型在实际应用当中与人类仍然存在较大的差距。这其中很大的一个原因就是，和机器阅读理解不同，人类在做阅读理解的时候，不仅仅借助所提供的文本内容，还会很好的去利用一些外部经验或者外部知识来辅助理解作答，这也正是机器阅读理解和人类阅读理解的一个巨大差异。因此，给机器阅读理解引入外部知识是一个很有意义也很具有挑战性的方向。

传统的方法中很大一部分是首先检索文本中某些词对应的外部知识，然后使用某种方法如知识图谱训练词向量的方法得到外部知识的一个词向量表示，然后再和原来文本的表示向量进行拼接或相加得到融合知识的文本表示，从而去增强机器阅读理解的性能。然而这种简单的融合方法可能会引入大量的噪声，反而导致效果不理想。近几年，随着预训练语言模型如 Bert 等的出现，有很多基于预训练语言模型设计的机器阅读理解模型，在很多机器阅读理解任务上都达到了不错的效果，然而使用预训练语言模型会导致文本中的实体词可能会被拆成几个片段，而大部分的知识都是实体级别的，这会导致机器阅读理解模型在融合外部知识的过程成引入更大的噪声，模型缺乏实体感知能力。

针对这个问题，本文提出了结合实体感知增强与外部知识的机器阅读理解模型 FSNER-net，该模型添加了命名实体识别的辅助任务进行联合训练，增强了机器阅读理解的实体感知能力，使得机器阅读理解更多的关注实体相关的信息，从而更好地融合实体相关的外部知识，最后在英文机器阅读理解数据集上的实验结果证实了我们方法的有效性。

本文其余部分组织如下：第二节介绍融合外部知识的机器阅读理解的相关工作；第三节介绍本文提出的结合实体感知增强与外部知识的机器阅读理解具体做法；第四节给出实验设置与结果分析；第五节简述结论及下一步工作展望。

2 相关工作

近几年，机器阅读理解发展迅猛。从 2015 年开始至 2018 年，陆续出现了很多机器阅读理解任务相关的数据集，像 2015 年发布的 CNN/DM^[2] 数据集、2016 年的 SQuAD^[3] 数据集、2016 年的 MS-MARCO^[4] 数据集、2017 年的 SearchQA^[5] 数据集以及 2017 年的 TriviaQA^[6]，这些数据集的发布极大的促进了端到端的机器阅读理解技术的快速发展，催生出了一系列的经典的机器阅读理解模型，像 Match-LSTM^[7]、BiDAF^[8]、AOA-Reader^[9]、DCN^[10]、R-Net^[11] 以及 QA-Net^[12]。这些端到端的机器阅读理解模型具有相似的模型架构，首先使用一个编码层分别获得问题和文章的向量表示，然后再利用注意力机制计算问题和文章的交互信息表示，最后再使用一个输出层例如常用的 Pointer-Network 来计算答案的起始位置和终止位置的概率分布。

而引入外部知识的机器阅读理近几年在国内外也出现了很多数据集以及相关工作。从 2018 年至今，出现了很多个高质量的需要借助外部知识的机器阅读理解的数据集，包括 ReCoRD^[13]、ARC^[14]、MCScripts^[15]、OpenBookQA^[16] 以及 CommonsenseQA^[17] 等。ReCoRD 可以看作是抽取式的机器阅读理解类型的数据集，而其他几个数据集都是多项选择式的机器阅读理解数据集，这些数据集都有一个共同点，那就是仅仅根据所提供的问题和文本，很难去回答问题的答案，需要借助一定的外部知识。这些数据集的出现极大地促进了引入外部知识的机器阅读理解研究的发展。

这里面的研究工作主要包含有引入辅助任务和改造融合外部知识结构两个方向的工作。Long^[18] 等就是引入了辅助任务。为了让基于深度学习的机器阅读理解模型能够利用外部知识，Long 等提出了一种新的辅助任务叫做稀有实体预测，该任务需要预测缺失的命名实体，类似与完形填空任务。然后借助层次化的循环神经网络引入了这些实体相关的描述信息。

改造融合外部知识结构的工作相对要多一点。Yang^[19] 等人为了让机器阅读理解引入外部

知识，直接传统的 LSTM 改造成 KBLSTM，从而同时编码文本和知识的语义信息。Sun^[20]等人利用 Key-Value 的 Memory Network 来找出相关的外部知识。所有可能用到的外部知识首先被选择放到知识库中的内存槽中作为键值对，键是用来和 query 做匹配，对应的值是相关知识的权重和表示。Weissenborn^[21]等提出了一种动态融合外部知识的方法，首先通过一个通用的阅读模块以文本的形式读入外部知识以及面向具体任务的文本内容，然后再根据具体任务来修正文本的内容表示，再去预测问题的答案。Mihaylov^[22]等针对完形填空式的阅读理解提出了一种将外部知识编码为 key-value 内存表示的 Knowledgeable Reader 网络，然后在推断出最后的答案之前将这些知识和文章的上下文表示进行融合，提高了机器阅读理解的效果。Pan^[23]等提出在科学问答数据集如 ARC 和 OpenBookQA 中利用信息检索技术引入无结构化的外部知识，并将该知识和 subject 相关的描述知识进行融合，辅助最后的答案选择。Yang^[24]等提出了一种 KT-NET 模块来融合外部知识，这些外部知识首先通过从知识图谱中检索得到，然后在 KT-NET 中通过 BiLinear 函数和 Attention 机制显示地融合这些外部知识，最后辅助机器阅读理解预测问题最终的答案。

本文受启发于这两种研究工作的思路，提出了结合实体感知增强与外部知识的机器阅读理解模型 FSNER-net，该模型通过添加命名实体识别的辅助任务，增强了模型的实体感知能力，使得机器阅读理解模型更多地关注实体信息，从而更好地融合实体相关的外部知识，进一步增强了模型结合外部知识的能力。

3 方法介绍

结合实体感知增强与外部知识的机器阅读理解模型 FSNER-net 包括命名实体识别的辅助任务和引入外部知识的机器阅读理解任务两大部分，为了训练这两个任务，需要两个关键的预处理：命名实体识别辅助任务数据构造和外部知识检索，接下来将先介绍这两部分，然后再着重介绍我们提出的模型 FSNER-net 的基本结构。

3.1 命名实体识别辅助任务数据构造

为了让机器阅读理解模型更多的关注实体相关的信息，增强机器阅读理解模型的实体感知能力，我们在机器阅读理解任务基础上添加了命名实体识别的辅助任务。然而机器阅读理解的原始数据集是没有命名实体识别标签数据的，因此我们在原始数据集上设定以下规则来构造命名实体识别数据，规则如下：

(1) 文章中的实体标注规则。根据机器阅读理解文章中的实体词信息，首先将文章中的词标注为实体，这里为了简便，只使用标签 ‘B’、‘I’、‘O’。如果是实体，使用标签 ‘B’ 和 ‘I’ 标注；如果不是实体，直接使用 ‘O’ 进行标注。例如 ‘Donald Trump’ 会被标注为 ‘B’ 和 ‘I’。并将标注的实体保存到一个列表里。

(2) 问题中的实体标注规则。根据列表里的实体词，来对问题文本中的实体进行标注。如果是实体，使用标签 ‘B’ 和 ‘I’ 标注；如果不是实体，直接使用 ‘O’ 进行标注。

通过这样的标注规则就得到了有命名实体识别标签的机器阅读理解数据集。

3.2 外部知识检索

为了在机器阅读理解中引入相关的外部知识，以增强机器阅读理解的性能，我们首先检索出机器阅读理解相关的外部知识。由于机器阅读理解数据集中很多的外部知识都是和实体相关的，所以我们采用 NELL¹知识库作为机器阅读理解的外部知识库。

NELL (Never-Ending Language Learner) 是卡内基梅隆大学开发的知识库，主要采用互联网挖掘的方法从 Web 自动抽取三元组知识，它的基本理念是：给定一个初始的本体（少量类和关系的定义）和少量样本，让机器能够通过自学习的方式不断的从 Web 学习和抽取

1 NELL 官网 <http://rtw.ml.cmu.edu/rtw/index.php?>

新的知识。目前 NELL 已经抽取将近 300 万条三元组知识。例如：special_events is a TV_show。

为了更好的检索出外部知识，针对 NELL 知识库的特点以及阅读理解数据集的特点，我们制定了以下处理规则来检索外部知识：

(1) 机器阅读理解数据集预处理：将机器阅读理解数据集里面的实体转换成字符串，去除标点符号并用_来替换空格；

(2) NELL 知识库预处理：处理 NELL 知识库中的实体，当实体中有数字的时候，去除前后的 n 符号；

(3) 检索规则：若机器阅读理解数据集中的实体包含一个词以上，检索知识的时候使用精确匹配；若只有一个词，先对该词进行词干提取，再部分匹配；

(4) 检索后处理：若机器阅读理解问题和文章中，如果一个实体 A 是实体 B 的前缀，则直接使用实体 B 代替实体 A。如使用实体 “Donald Trump” 代替实体 “Trump”。

根据该方法，我们检索出了文本中的一些实体的相关知识，抽选出一部分的结果如表 3-1 所示。

表 3-1 实体知识检索结果表

实体名称	检索知识列表
Anthony Watson	athlete
England	organization; island; trainstation; geopoliticalorganization
cheek	muscle; bodypart
far from the madding crowd	book
department of justice	governmentorganization
donald trump	celebrity; ceo; male

可以看到，这些检索出来的知识是对原始文本词的语义补充，将这些知识作为机器阅读理解数据的外部知识库。

为了更好的得到知识词向量的初步表示，我们采用经典的 TransE^[25]模型来预训练知识图谱的词向量，这里我们直接使用了清华开源的 OpenKE²工具，将训练完的词向量作为外部知识的初始化表示。

3.3 结合实体感知增强与外部知识的阅读理解模型

我们将提出的结合实体感知增强与外部知识的机器阅读理解模型简称为 FSNER-net，模型的结构图如图 3-1 所示，该模型主要包含 5 个部分，分别是：(1)BERT 编码层 (BERT Encoding layer)，用来计算问题和文章的深度文本表示；(2)知识融合层 (Fusion Knowledge layer)，用来融合文本和知识的表示；(3)自注意力层 (Self-Attention layer)，用来计算文本知识表示的自注意力机制；(4)序列标注层 (Sequence Tagging layer)，用来对文本进行序列标注。(5)输出层 (Output layer)，用来预测答案的起始位置和终止位置；接下来，将详细说明模型 FSNER-net 每一部分具体是怎么做的。

(1) BERT 编码层

为了更好的得到问题和文章的深度文本表示，我们使用 BERT 编码层编码问题和文章。假设文章表示为 $P = \{p_i\}_{i=1}^m$ ，问题表示为 $Q = \{q_j\}_{j=1}^n$ ，为了方便输入 BERT，将它们拼接成如下形式：

$$S = [< CLS >, Q, < SEP >, P, < SEP >] \quad (3-1)$$

这种输入表示经过 BERT，得到最后的隐层表示 $\{h_i^L\}_{i=1}^{m+n+3}$ ，其中 L 表示 BERT 层数。

2 清华知识图谱 OpenKE 工具箱: <https://github.com/thunlp/OpenKE>

(2) 知识融合层

得到了问题和文章的深度文本表示之后，为了得到更好的融合外部知识的文本表示，将知识表示和文本表示通过知识融合层来计算融合知识的文本表示。经过 BERT 之后的文本表示的每个词可以看作是一个 token，对于每个 token s_i ，它的 BERT 表示 $h_i^L \in R^{d_1}$ 。在训练之前，会按照前面提到的检索方法给每个 token s_i 检索到一组相关的 KB concepts 集合称为 $C(s_i)$ ，而且集合中的每个 concept c_j 都有一个预训练的 KB embedding $c_j \in R^{d_2}$ 。

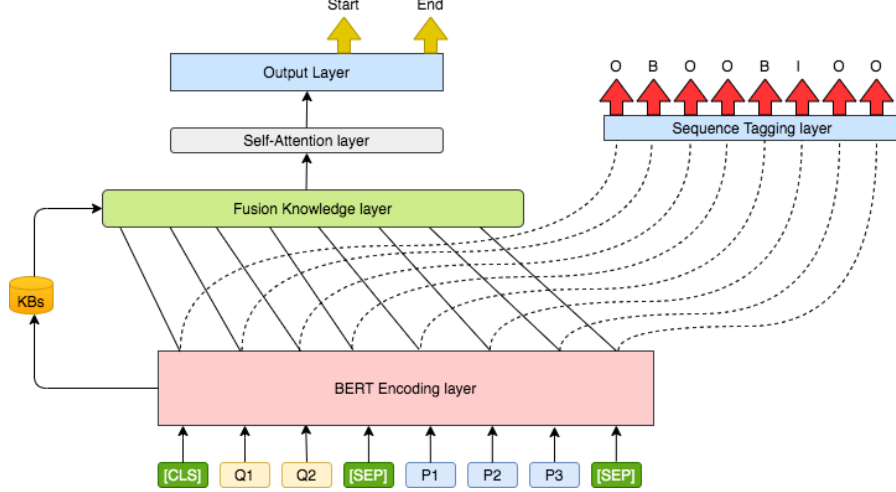


图 3-1 FSNER-net 模型结构图

为了更好的选择性地融合这些外部知识向量，这里采用了注意力机制。对于每个 token s_i ，为了衡量 c_j 与 s_i 的相关程度，采用 bilinear 操作来计算这个注意力权重，计算公式如下：

$$\alpha_{ij} \propto \exp(c_j^T W h_i^L) \quad (3-2)$$

其中 $W \in R^{d_2 \times d_1}$ 。另外，由于不是每个 token s_i 都会有知识需要引入，因此，这里引入了一个空知识向量 $\bar{c} \in R^{d_2}$ ，并用同样的方式计算注意力权重为：

$$\beta_i \propto \exp(\bar{c}^T W h_i^L) \quad (3-3)$$

对于 token s_i ，对齐检索的知识集合得到融合外部知识的知识状态向量 k_i ，

$$k_i = \sum_j \alpha_{ij} c_j + \beta_i \bar{c} \quad (3-4)$$

其中 $\sum_j \alpha_{ij} + \beta_i = 1$ 。

最后拼接 BERT 表示 h_i^L 和知识状态向量 k_i 得到 $u_i = [h_i^L, k_i] \in R^{d_1+d_2}$ ， u_i 就具有了上下文信息和相关外部知识信息。

(3) 自注意力层

为了得到更深度的融合知识文本的交互信息表示，采用 Self-Attention 机制来进一步计算上下文的信息表示和相关外部知识信息的交互信息表示，而且这里同时使用了直接的交互信息表示和间接的交互信息表示。

在直接的交互信息表示中，对于 token s_i 和 s_j ，它们的外部知识信息表示分别为 u_i 和 u_j ，这里首先拼接 u_i 与 u_j 的差、和以及它们之间的哈达马乘积，然后再过一个线性层来计算相似矩阵，

$$r_{ij} = w^T [u_i + u_j, u_i - u_j, u_i \odot u_j] \quad (3-5)$$

其中 $w \in R^{3d_1+3d_2}$ 是可训练的参数，这样得到一个矩阵 R ，然后对矩阵 R 做一个行的 softmax，得到自注意力的权重矩阵 A ，然后对于每个 token s_i 会得到一个计算注意力之后的 vector v_i ，

$$a_{ij} = \frac{\exp(r_{ij})}{\sum_j \exp(r_{ij})} \quad (3-6)$$

$$v_i = \sum_j a_{ij} u_j \quad (3-7)$$

间接的交互信息表示中, 对原来的自注意力的权重矩阵 A 进行一个自乘, 然后得到每个 token s_i 的另外一个信息表示向量 \bar{v}_i ,

$$\bar{A} = A^2 \quad (3-8)$$

$$\bar{v}_i = \sum_j \bar{a}_{ij} u_j \quad (3-9)$$

最后将结果拼接起来得到 $o_i = [u_i, v_i, u_i - v_i, u_i \odot v_i, \bar{v}_i, u_i - \bar{v}_i] \in R^{6d_1+6d_2}$ 。

(4) 序列标注层。

为了让机器阅读理解模型更多的关注实体相关的信息, 增强模型的实体感知能力, 这里引入了命名实体识别的辅助任务, 它与机器阅读理解部分共享 BERT 编码层, 两个任务联合训练。具体的, 这一层用来对文本进行序列标注, 得到问题和文章的深度表示之后过一个线性层, 计算每个 token s_i 标注为标签 c_k 的概率, 使用交叉熵损失函数, 为:

$$p_{c_k} = \text{softmax}(w^T h_i^L) \quad (3-10)$$

这一部分的目标函数使用交叉熵损失函数,

$$\mathcal{L}_1 = -\frac{1}{N} \sum_{j=1}^N \sum_{k=1}^C (y_k \log p_{c_k}) \quad (3-11)$$

(5) 输出层

这一层用来计算预测答案的起始位置和终止位置的概率分布。将自注意力层的输出结果过一个线性层, 然后再过一个标准的 softmax, 计算预测答案的边界, 即每个 token s_i 作为答案片段的起始位置的概率和终止位置的概率, 即

$$p_i^1 = \frac{\exp(w_1^T o_i)}{\sum_j \exp(w_1^T o_j)}, \quad p_i^2 = \frac{\exp(w_2^T o_i)}{\sum_j \exp(w_2^T o_j)} \quad (3-12)$$

其中 $w_1, w_2 \in R^{6d_1+6d_2}$ 是可训练的参数。

这一部分模型训练的目标函数采用最大似然, 为

$$\mathcal{L}_2 = -\frac{1}{N} \sum_{j=1}^N (\log p_{y_j^1}^1 + \log p_{y_j^2}^2) \quad (3-13)$$

在推断的时候, 将概率乘积最大的 $p_a^1 p_b^2$ 所对应的片段 (a, b) ($a < b$) 作为最后预测的答案。

由于是联合训练命名实体识别辅助任务和融合外部知识的机器阅读理解任务, 所以本模型最终的目标函数包含了两部分, 一部分是机器阅读理解预测起始位置和终止位置的极大似然, 另外一部分是命名实体识别部分标注的交叉熵损失, 总的目标函数是:

$$\mathcal{L} = -\frac{1}{N} \sum_{j=1}^N (\log p_{y_j^1}^1 + \log p_{y_j^2}^2) - \frac{1}{N} \sum_{j=1}^N \sum_{k=1}^C (y_k \log p_{c_k}) \quad (3-14)$$

4 实验

4.1 数据来源

我们在英文的机器阅读理解 ReCoRD 数据集进行实验。ReCoRD 数据集任务需要借助外部知识推理, 该外部知识推理大致有 5 种类型: 意译 (3%)、部分线索 (10%)、多个句子的推理 (6%)、常识推理 (75%)、歧义 (6%)。这其中常识推理再细分包括: 概念知识 (49.3%)、

因果推理 (32.0%)、通俗心理学 (28.0%)、社会规范和空间推理等 (12.0%)。它的形式类似于 SQuAD，同样是给一篇文章和一个问题，要求从文章中找出该问题的答案，答案是文章中的一个片段。该数据集包含训练集 100K，验证集 10K，测试集合 10K。

4.2 评价指标

由于将 ReCoRD 数据集任务为抽取式的机器阅读理解，所以采用 EM 值 (Exact Match) 和 F1 值 (F1-score) 的评价指标来进行综合评价。

(1) F1 值的计算方法为：计算预测答案与原始真实答案字符之间的重叠，根据字符重叠的数目与原始真实答案的字符数目计算召回率，字符重叠的数目与预测出的所有字符数目计算准确率，F1 值计算公式如下：

$$F_1 = \frac{2 * \text{准确率} * \text{召回率}}{\text{准确率} + \text{召回率}} \quad (4-1)$$

(2) EM 值的计算方法：表示预测答案与原始真实答案是否完全匹配，如果完全匹配则为 1，否则为 0。

4.3 实验结果及分析

4.3.1 机器阅读理解实验结果

我们在 ReCoRD 数据集上进行实验，实验结果如表 4-1 所示。

表 4-1 FSNER-net 及其相关模型实验结果表

模型	Dev		Test	
	EM	F1	EM	F1
QANet ^[12]	35.38	36.75	36.51	37.79
DocQA ^[26] w/o ELMO	36.59	37.89	38.52	39.76
DocQA ^[26] w/ ELMO	44.13	45.39	45.44	46.65
BERT ^[1] (base)	54.03	55.99	54.04	56.07
BERT ^[1] (large)	64.28	66.60	—	—
BERT (large-wwm)	71.59	73.69	—	—
FSNER-net _{bert-large-wwm} +NELL (ours)	73.17	75.05	73.57	75.63

表中 QANet^[12] 是机器阅读理解中比较有名的模型，它的编码器仅由卷积和自注意力机制组成，是相比于传统问答经典架构的一种新型架构，在 SQuAD 数据集上能达到很不错的效果，然而在 ReCoRD 数据集上的实验结果欠佳，这也说明了 ReCoRD 数据集的难度。DocQA 是针对阅读理解中有的文章无法回答问题而设计的模型，ELMo^[28] 是基于特征的预训练语言模型，同样是利用了大量的无标注的文本预训练。可以看到单凭 DocQA 模型达到的效果也是差强人意，而加上 ELMo，效果提升幅度较大，这也说明了基于特征的预训练语言模型的有效性。

我们提出的 FSNER-net 模型是基于预训练语言模型而设计的，为了选择更强的预训练语言模型，首先使用单纯的预训练语言模型进行实验。使用预训练语言模型 BERT 之后，可以看到效果相比于传统的阅读理解方法有大幅度提升，而且 BERT-large 的效果相比于 BERT-base 的效果也有大幅度地提升，这也说明了 BERT 等预训练语言模型借助优秀的深度学习架构，能够很好的利用大量的无标注的文本，而且这些大量的无标注的文本中已经包含了很多的外部知识，因此预训练模型在该任务上表现不错。另外，我们还使用更换了原始 BERT 的 mask 机制的全词 mask (whole word masking) 策略预训练的 BERT 语言模型，相比于原始的 BERT，提升效果也很明显，这也表明 ReCoRD 数据集中的文本需要借助很多实体相关的外部知识，而预训练语言模型全词 mask 策略更有利于这些实体信息的学习，从而能够让预训练语言模型学习到更多实体相关的知识信息。

因此，我们提出的 FSNER-net 机器阅读理解模型以 BERT-large-wwm 为基础，该模型不仅添加了命名实体识别辅助任务，而且从 NELL 知识库中有效地引入了外部知识。可以看到，FSNER-net 在开发集上的效果已经远远超过了传统的 QANet、DocQA 包括 BERT 等方法，说明了我们方法的优越性。另外相比于强大预训练语言模型 BERT-large-wwm，提升效果也很显著，在开发集上 EM 值提升 1.58，F1 值提升 1.36。这也说明即使是强大的预训练语言模型，可能借助优秀的深度学习架构、适合的 mask 机制，已经在大量的无标注的文本中学习到了大量的外部知识，但是借助我们提出的结合实体感知增强和外部知识的机器阅读理解方法还能够进一步提升机器阅读理解的效果，说明联合命名实体识别的辅助任务增强了机器阅读理解的实体感知能力，使得机器阅读理解更多地关注实体相关的信息，这将更有助于实体相关的外部知识的引入。

4.3.2 模型消融实验

为了更好的分析 FSNER-net 模型各个部分的作用，我们还进行了模型消融实验，来查看模型的每个部分的贡献，实验结果表 4-2 所示：

表 4-2 FSNER-NET 模型消融实验结果表

模型	Dev	
	EM	F1
FSNER-netbert-large-wwm+NELL	73.17	75.05
w/o NELL	71.59	73.69
w/o NER task	72.31	74.42
w/o self-attention	72.85	74.95

从模型消融的实验结果可以看出 FSNER-net 模型各个部分都起到了作用。可以看到，NELL 知识库的引入极大地增强了机器阅读理解的性能，相比于引入 NELL 外部知识，F1 值大约提升 1.36 个点，这说明了 NELL 知识库中的实体信息对于机器阅读理解是有很大帮助的，也说明了我们显示地给机器阅读理解引入外部知识方法的有效性；另外，即使已经引入了外部知识，联合命名实体识别的辅助任务进一步提高了机器阅读理解的效果，这说明了联合命名实体识别的辅助任务的确有助于增强机器阅读理解的实体感知能力，使得机器阅读理解模型能够更多的关注实体信息，从而更有利于机器阅读理解模型融合实体相关的外部知识信息。另外，FSNER-NET 的中的 self-attention 也能有少许提升，这很可能使因为 self-attention 能够进一步得到融合文本的交互信息表示，也说明传统机器阅读理解架构的有效性。总之，我们提出的 FSNER-net，不仅给机器阅读理解有效地引入了外部知识，而且联合命名实体识别的辅助任务提高了机器阅读理解的实体感知能力，进一步提高了机器阅读理解模型融合外部知识的能力，最终在效果上远超传统的方法。

5 结论及下一步展望

本文提出了结合实体感知增强和外部知识的机器阅读理解模型 FSNER-net，既通过有效的方法融合了外部知识，又创造性地联合命名实体识别任务提高了机器阅读理解模型的实体感知能力，从而进一步提高了机器阅读理解模型引入实体相关的外部知识的能力，通过实验对比，最终在效果上远超目前其他的方法。下一步我们的研究将尝试使用更大的预训练语言模型 XLNET^[29]、ALBERT^[30]等为基础来设计模型，或者在预训练语言模型预训练的过程中直接添加实体相关的任务如命名实体识别任务，来直接增强预训练语言模型的实体感知能力。

感谢

感谢张宇老师对我的指导和帮助，感谢 QA 组师兄们对我的帮助。

引用

[1] Devlin, Jacob, 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, Volume 1 (Long and Short Papers). 2019.
- [2] Hermann, Karl Moritz, et al. "Teaching machines to read and comprehend." Advances in neural information processing systems. 2015.
- [3] Rajpurkar, Pranav, et al. "SQuAD: 100,000+ Questions for Machine Comprehension of Text." Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016.
- [4] Nguyen, Tri, et al. "MS MARCO: a human-generated machine reading comprehension dataset." (2016).
- [5] Dunn, Matthew, et al. "Searchqa: A new q&a dataset augmented with context from a search engine." arXiv preprint arXiv:1704.05179 (2017).
- [6] Hermann, Karl Moritz, et al. "Teaching machines to read and comprehend." Advances in neural information processing systems. 2015.
- [7] Wang, Shuohang, and Jing Jiang. "Machine Comprehension Using Match-LSTM and Answer Pointer." (2016).
- [8] Seo, Minjoon, et al. "Bidirectional Attention Flow for Machine Comprehension." (2016).
- [9] Cui, Yiming, et al. "Attention-over-Attention Neural Networks for Reading Comprehension." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017.
- [10] Xiong, Caiming, Victor Zhong, and Richard Socher. "Dynamic coattention networks for question answering." arXiv preprint arXiv:1611.01604 (2016).
- [11] Wang, Wenhui, et al. "Gated self-matching networks for reading comprehension and question answering." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017.
- [12] Yu, Adams Wei, et al. "QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension." (2018).
- [13] Zhang, Sheng, et al. "Record: Bridging the gap between human and machine commonsense reading comprehension." arXiv preprint arXiv:1810.12885 (2018).
- [14] Clark, Peter, et al. "Think you have solved question answering? try arc, the ai2 reasoning challenge." arXiv preprint arXiv:1803.05457 (2018).
- [15] Ostermann, Simon, et al. "Semeval-2018 task 11: Machine comprehension using commonsense knowledge." Proceedings of the 12th International Workshop on semantic evaluation. 2018.
- [16] Mihaylov, Todor, et al. "Can a Suit of Armor Conduct Electricity? A New Dataset for Open Book Question Answering." Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018.
- [17] Talmor, Alon, et al. "CommonsenseQA: A Question Answering Challenge Targeting Commonsense Knowledge." Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019.
- [18] Long, Teng, et al. "World knowledge for reading comprehension: Rare entity prediction with hierarchical lstms using external descriptions." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

-
- [19] Yang, Bishan, and Tom Mitchell. "Leveraging Knowledge Bases in LSTMs for Improving Machine Reading." Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017.
- [20] Sun, Yibo, et al. "Knowledge based machine reading comprehension." arXiv preprint arXiv:1809.04267 (2018).
- [21] Weissenborn, Dirk, Tomas Kocisky, and Chris Dyer. "Dynamic Integration of Background Knowledge in Neural NLU Systems." (2018).
- [22] Mihaylov, Todor, and Anette Frank. "Knowledgeable Reader: Enhancing Cloze-Style Reading Comprehension with External Commonsense Knowledge." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018.
- [23] Pan, Xiaoman, et al. "Improving question answering with external knowledge." arXiv preprint arXiv:1902.00993 (2019).
- [24] Yang, An, et al. "Enhancing pre-trained language representations with rich knowledge for machine reading comprehension." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019.
- [25] Bordes, Antoine, et al. "Translating embeddings for modeling multi-relational data." Advances in neural information processing systems. 2013.
- [26] Rajpurkar, Pranav, Robin Jia, and Percy Liang. "Know What You Don't Know: Unanswerable Questions for SQuAD." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2018.
- [27] Clark, Christopher, and Matt Gardner. "Simple and Effective Multi-Paragraph Reading Comprehension." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018.
- [28] Peters, Matthew E., et al. "Deep contextualized word representations." Proceedings of NAACL-HLT. 2018.
- [29] Yang, Zhilin, et al. "Xlnet: Generalized autoregressive pretraining for language understanding." Advances in neural information processing systems. 2019.
- [30] Lan, Zhenzhong, et al. "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations." International Conference on Learning Representations. 2019.

作者联系方式: 乐远 黑龙江省哈尔滨市南岗区哈工大一校区新技术楼 602 室 150001
13069875982 yle@ir.hit.edu.cn