

北京交通大学

M3 阅读笔记



班 级： 通信 1907
姓 名： 赵司祺
学 号： 19211449
任课教师： 陈一帅

一、 文献信息

作者： Venkat Arun, Mina Tahmasbi Arashloo, Ahmed Saeed, Mohammad Alizadeh, Hari Balakrishnan

论文题目： Commonsense Reasoning with Implicit Knowledge in Natural Language (自然语言中隐含知识的常识推理)

二、 问题意义

对于人工智能来说，要对日常的人类活动进行推理，它需要具备常识。因此，常识的获取和推理被认为是人工智能早期的关键研究挑战。近年来，一些自然语言物理交互任务被提出，在这些问答环节中常识推理非常重要。实现这一点的两种常见方法是：

- 1、 使用认知图中结构良好的常识；
- 2、 使用逐渐强大的 transformer 语言模型。(Transformer 是一个利用注意力机制来提高模型训练速度的模型。)

在方法 1 中在获取和表示常识方面很具有挑战性，方法 2 需要大量的资源。所以在这项工作中，采取了折中的策略，即使用较小的语言模型和相对较小但有针对性的自然语言文本语料库。这种方法的优点是资源密集度较低，同时可以使用非结构化文本语料库。

本篇文章中定义了不同的非结构化常识知识源，探索了三种知识整合策略，并提出了四种与最先进的方法相比具有竞争力的方法来推理隐含常识。

三、 思路方法

为了研究 transformer 的常识推理能力，选择 aNLI (判断人工智能系统进行诱因推理的潜力，以形成对给定观测集的可能解释)、PIQA (评估 AI 系统的物理推理能力)、SIQA (关于社会互动推理及其陈述的社会影响的实例集合) 这三个数据集来评估本篇文章中研究的模型，每个数据集都有不同的常识知识。图 1 显示了每个数据集的示例，以及我们检索到的常识句子。

Abductive NLI	Social IQA	Physical IQA
<p>Obs1: Jim was working on a project. ✓ Jim found he was missing an item. ✗ Jim needed a certain animal for it. Obs2: Luckily, he found it on a nearby shelf Knowledge: Peyton eventually found it before Peyton needed to determine that something is missing. Kendall never found it, as a result Kendall wants to lodge a missing complaint.</p>	<p>Context: Remy was an expert fisherman and was on the water with Kai. Remy baited Kai's hook. Question: What will Remy want to do next? ✓ cast the line ✗ put the boat in the water ✗ invite Kai out on the boat Knowledge: Alex baits Pat's hook as a result others want to cast their line.</p>	<p>Goal: When doing sit-ups: ✓ place your tongue in the roof of your mouth. It will stop you from straining your neck. ✗ place your elbow in the roof of your mouth. It will stop you from straining your neck. Knowledge: How to Do Superbrain Yoga. Place your tongue on the roof of your mouth.</p>

图 1 三个数据集以及检索到的知识的示例

然后对于常识性知识来源进行了分类，使用两个 transformer 模型，深入分析 transformer 在三个不同的常识 QA 任务中使用隐式知识进行常识推理的能力。

接着文章介绍了知识注入的四个模型，分别是合并多个字符串（Concat）、并行最大（Parallel-Max）、简单求和（Simple Sum）以及加权求和（Simple Sum），这些方法适用于多个常识推理任务，与预先训练的 transformer 相比，性能提高了 2-9%。前三个，合并多个字符串（Concat）、并行最大（Parallel-Max）和简单求和（Simple Sum）作为更强的基线，使用与我们提出的加权和模型相同的隐式知识。

作者在提出了一个知识注入方法的实验框架后进行了实验，并观察到与强基线相比有 2-9% 的显著改善。

Dataset	Strategy	BERT				RoBERTa			
		Concat	Max	Sim-Sum	Wtd-Sum	Concat	Max	Sim-Sum	Wtd-Sum
aNLI	OPENBOOK	73.9± 0.8	73.7± 0.1	73.5± 0.7	73.3± 1.0	83.9± 0.5	80.8± 0.9	81.7± 0.6	84.4± 0.4
	REVISION	72.7± 0.3	N/A	N/A	N/A	82.4	N/A	N/A	N/A
	REVISION & OPENBOOK	74.4± 0.2	74.3± 0.1	74.0± 0.9	75.1± 0.4	84.2± 0.7	81.4± 0.8	82.6± 0.6	86.7± 0.6
PIQA	OPENBOOK	67.8± 0.4	72.4± 0.6	72.6± 1.2	72.5± 0.1	74.8± 0.5	75.2± 0.9	75.6± 0.7	77.1± 0.2
	REVISION	74.5± 0.3	N/A	N/A	N/A	75.2± 0.8	N/A	N/A	N/A
	REVISION & OPENBOOK	67.7± 0.1	73.8± 0.8	76.8± 0.5	76.8± 0.3	75.4± 0.7	76.2± 0.8	76.8± 0.4	80.2± 0.6
SIQA	OPENBOOK	70.1± 0.8	67.8± 0.1	70.0± 0.7	70.2± 0.4	76.5± 0.7	77.2± 0.6	77.4± 0.2	78.3± 0.5
	REVISION	69.5± 0.9	N/A	N/A	N/A	76.8± 0.3	N/A	N/A	N/A
	REVISION & OPENBOOK	68.8± 0.4	66.6± 0.4	68.9± 0.1	69.3± 0.6	78.2± 0.3	77.4± 0.9	76.7± 0.5	79.5± 0.9

表 1 四种模型（Concat、Max、Simple sum、Weighted sum）的验证集精度（%）

Models/ Accuracy	aNLI		PIQA		SIQA	
	Val	Test	Val	Test	Val	Test
BERT	67.36	66.75	68.08	69.23	64.88	64.50
GPT-2 XL	N/A	N/A	70.20	69.50	47.50	45.30
RoBERTa	85.05	83.91	76.28	76.80	77.85	76.74
RoBERTa 5 Ensemble	N/A	83.22	N/A	79.66	N/A	78.68
L2R ² [2020]	N/A	86.81	N/A	N/A	N/A	N/A
KagNet [2019]	N/A	N/A	N/A	N/A	65.05	64.59
GBR [2020]	N/A	N/A	N/A	N/A	75.64	76.25
UnifiedQA T5 11B [2020]	N/A	80.04	N/A	89.50	N/A	79.75
Ours: BERT + WS	74.60	74.96	76.82	72.28	70.21	67.22
Ours: RoBERTa + WS	85.90	84.18	80.20	78.24	79.53	78.00

表 2 与当前最佳方法相比，采用修订版和 Openbook 策略的加权和模型的性能

Model	Knowledge Source	aNLI	PIQA	SIQA
BERT	Directly/Partially Derived	75.1± 0.4	N/A	70.2± 0.4
	TrainOnly Directly/Partially	74.6± 0.8	N/A	69.8± 0.7
	Related Knowledge	73.2± 0.5	76.8± 0.3	68.6± 0.5
RoBERTa	Directly/Partially Derived	86.7± 0.6	N/A	79.5± 0.9
	TrainOnly Directly/Partially	85.9± 0.8	N/A	78.9± 1.2
	Related Knowledge	85.0± 1.1	80.2± 0.6	77.4± 0.8

表 3 不同知识源类型对加权和知识注入模型的影响

表 1、2 和 3 总结了在三个数据集上的结果。BERT 和 RoBERTa 基线验证和隐藏测试分数见表 2。以自然语言形式添加知识在统计学上显著提高了所有数据集的 QA 准确性，高于基线 BERT， $p \leq 0.05$ 基于 Wilson 得分区间。这包括从相关知识源中检索知识，如表 2 和表 3 所示。合并多个字符串（concat）模式的知识注入比基线 BERT 提高了 1-6%，加权求和模型（加权求和）进一步提高了 2-4%。在表 2 中可以观察到，在 aNLI 中，加权和模型比 T5 好 4.1%，在 SIQA 中，参数数量减少了 30 倍，差距缩小到 1.75%。它还超越了基于复杂图

形的方法，如 GBR 和 KagNet。此外，文章中使用的方法更具样本效率，使用自然语言表达的隐性知识能够弥补与超大 transformer 之间的差距。

四、 实验结论

在这项工作中，文章作者全面研究了 transformer 利用自然语言表达的隐性知识进行推理的能力。作者提出了一个知识注入方法的实验框架，并观察到与强基线相比有 2-9% 的显著改善。这种方法使用较少的样本和参数进行训练，使用大量预先训练的语言模型进行竞争，并超越基于复杂图形的方法。此外这种方法足够通用，可以应用于其他知识密集型任务和语言。这种方法缩小了小型和大型预培训 transformer 之间的差距。文章还对不同的组成部分进行了批判性分析，发现即使有足够的知识，transformer 在 30-50% 的时间内仍然无法回答问题，日后还需要更好的方法来使用隐含知识进行推理。

五、 启发思考

当今人工智能领域飞速发展，由于日常生活中使用自然语言已成为理所当然的事情，导致我从来没有想过，世界上有那么多种语言，每一种语言都有其特殊的内在逻辑，所以我们平常所说的话人工智能是不是能正确理解含义，并且给出正确解答？在许多情况下，人工智能需要进行推理得出答案，推理所必需的常识可能以隐含的方式出现在文本中，所需的常识知识可能无法在非结构化知识库中明确获得，并且检索到的知识和上下文问题答案的高度词汇重叠并不意味着它可以用来正确回答。

这篇文章提出了解决这类问题的一种让人耳目一新的方法：提出了一个知识注入方法的实验框架，使用大量预先训练的语言模型进行竞争，并超越基于复杂图形的方法。

这是我第一次阅读这样一篇如此长篇幅的全英文论文，翻译软件对于大段的专业名词的翻译欠佳，有时甚至完全读不通，于是只能通过逐个查询生词的方法慢慢阅读，慢慢体会论文所表达的含义。从本次的阅读英文论文的经历中，我知道了要想没有损耗的接收论文作者所表达阐述的观点和结论，最好的方法就是阅读英文原文，而不是通过翻译软件进行大段的翻译，因此，扎实的英语基础对于本专业的学习也是大有裨益的。

我会在后续的学习中加强对于英语的学习，以及对于数学基础的巩固，对于英文文献不再惧怕，在专业课的学习过程中多多查找文献资料来丰富学习成果，争取将专业知识学得更好。