

基于时态知识图的问答

阿普尔·萨克塞纳
印度科学研究所

班加罗尔

apoorvsaxena@iisc.ac.in

苏门·查克拉巴蒂
印度理工学院

孟买

soumen@cse.iitb.ac.in

帕萨·塔鲁克达尔

谷歌研究

印度

帕莎@google.com

抽象的

时态知识图 (时态 KG)

通过提供时间范围 (例如开始和结束时间)来扩展常规知识图

在 KG 的每条边上。虽然问题 KG (KGQA) 的转向已收到一些

研究界、QA 的关注

时间 KG (Temporal KGQA) 是

相对未开发的领域。缺乏广泛覆盖的数据集是限制该领域进展的另一个因素。我们解决这个问题

通过提出CRONQUESTIONS来挑战，已知最大的时态 KGQA 数据集，清晰地分层为结构复杂性的桶。CRONQUESTIONS 扩展了唯一的已知先前数据集的 340 倍。

我们发现各种最先进的 KGQA

方法在这个新数据集上远远达不到预期的性能。作为回应，

我们还提出了CRONKGQA,这是一种基于变压器的解决方案,它利用了最新的进展

时间 KG 嵌入,性能优于所有基线,准确率比次优提高 120%

执行方法。通过大量的实验,我们详细了解了 CRONKGQA 的工作原理以及情况

出现显著的进一步改进

可能的。除了数据集之外,我们还发布了我们的代码。

1 简介

时态知识图 (Temporal KG)是

多关系图,其中每条边都与持续时间相关联。这与

没有时间注释的常规 KG。

例如,一个常规 KG 可能包含一个事实

例如 (巴拉克·奥巴马,担任职务,总统

美国),而临时 KG 将包含

以及开始和结束时间 - (巴拉克·奥巴马,举行

职位,美国总统,2008 年,2016 年)。边缘

可能与一组不连续的

时间间隔也是如此。这些时间范围

事实可以自动估计 (Taluk dar et al., 2012)或用户贡献。

好几个这样的

文献中已经提出了时间 KG,

其中重点是 KG 完成 (Dasgupta

等人。2018;加西亚-杜兰等人。2018;利太鲁和

施罗德2013;拉克鲁瓦等人。2020;贾恩等人。2020)。

知识图问答 (KGQA)的任务是使用知识图谱作为知识库来回答自然语言问题。这是

与基于阅读理解的问答相反,在这种问答中,问题通常伴随着上下文 (例如,文本段落)并且

答案要么是多项选择之一 (Rajpurkar et al., 2016),要么是来自

上下文 (Yang 等人, 2018)。在KGQA中,答案通常是KG 中的一个实体 (节点),而

回答问题所需的推理是

基于单一事实 (Bordes et al., 2015),多跳

(Yih 等人,2015 年; Zhang 等人,2017 年)或基于连接/比较的推理 (Talmor 和 Berant, 2018 年)。时间 KGQA

迈出了这一步

进一步其中:

- 1.底层KG是一个Temporal KG。
2. 答案是实体或持续时间。
3. 可能需要复杂的时间推理。

KG 嵌入是 KG 中实体和关系的低维密集向量表示。

文献中提出了几种嵌入知识图谱的方法 (Bordes et al. 2013, Trouillon

等人。2016, Vashishth 等人。2020)。这些嵌入最初是为了 KG 的任务而提出的

完成,即预测缺失的边缘

KG,因为大多数现实世界的 KG 都是不完整的。

然而最近它们也被应用到

KGQA 的任务已被证明

提高完整和不完整 KG的设置的性能 (Saxena et al. 2020; Sun

等人。2020)。

数据集	公斤	显事实	问题类型			# 问题
			多实体多关系时态			
简单问题	免费基地				0%	108k
目标是质量检查	梅塔QA 公斤			✓	0%	40万
网络问题	FreeBase	✓		✓	<16%	5,810
复杂网络问题 FreeBase				✓	-	35k
临时问题	自由基地			✓	100%	1,271
CRONQUESTIONS (我们的)	WikiData		✓✓✓	✓	100%	41万

表 1:KGQA 数据集比较。统计有关 WebQuestions 的时间问题的百分比

来自贾等人。(2018a)。我们没有 ComplexWebQuestions 的明确数量的时间问题,但是由于它是使用 WebQuestions 中的问题自动构建的,因此我们预计百分比相似到网络问题 (16%)。详细内容请参见2.1节。

时间 KG 嵌入是另一个即将推出的实体、关系和时间戳的区域。时间 KG 嵌入在低维中向量空间 (Dasgupta et al. 2018, Lacroix et al. 2018) 2020,贾恩等人。2020,戈埃尔等人。2019)。这里也,到目前为止主要的应用是时间KG完成。在我们的工作中,我们调查是否时间 KG Embeddings 可以应用于时态 KGQA 的任务,以及它们与非时态嵌入或现成的相比如何

没有任何 KG 嵌入的方法。在本文中,我们提出了CRONQUESTIONS,时间 KGQA 的新数据集。CRONQUESTIONS由时间知识图谱和附带的自然语言问题组成。有

创建此项目时的三个主要指导原则

数据集:

1.关联的 KG 必须提供时间符号。

2.问题必须涉及时间因素推理。

3.标记实例数量一定要大足以用于训练模型,

而不是仅仅为了评价。在上述原则的指导下,我们提出了由 125k 的 Temporal KG 组成的数据集实体和 328k 事实,以及一组 410k 需要时间的自然语言问题

推理。在这个新数据集上,我们应用基于的方法

仅在深度语言模型 (LM) 上,例如 T5 (Raffel 等人, 2020)、BERT (Devlin 等人, 2019)、和 KnowBERT (Peters 等人, 2019),以及混合 LM+KG 嵌入方法,例如实体作为专家 (Fevry 等人, 2020)和 EmbedKGQA (Saxena 等人, 2020)。我们发现这些基线不适合时间推理。作为回应,我们提出了CRONKGQA,

EmbedKGQA 的增强,其性能优于

所有问题类型的基线。克朗克质量保证

在简单的时间上达到非常高的精度

推理问题,但到了时候就达不到要求

需要更复杂推理的问题。

因此,尽管我们获得了有希望的早期结果,CRONQUESTIONS留下了足够的改进空间复杂的时间 KGQA。我们的源代码CRONQUESTIONS数据集可以在以下位置找到<https://github.com/apoorvumang/CronKGQA>。

2 相关工作

2.1 时间 QA 数据集

已经提出了几个 KGQA 数据集

文献中 (表1)。在 SimpleQuestions (Bordes et al., 2015)中,人们只需要提取一个其实还是从KG来回答一个问题。元质量保证 (Zhang 等人, 2017)和 WebQuestionsSP (Yih et al., 2015)需要多跳推理,其中必须遍历 KG 中的多条边来得出答案。ComplexWebQuestions (Tal mor 和 Berant, 2018)包含多跳和连接/比较类型的问题。然而,这些都不是针对时间推理的,并且

它们所基于的 KG 是非时间性的。主要研究了时态 QA 数据集

在阅读理解领域。一个这样的数据集是 TORQUE (Ning et al., 2020),其中

系统被给出一个问题以及一些上下文

(一段文字)并被要求回答多个

有五个选项的选择题。这与 KGQA 形成对比,KGQA 没有上下文,并且答案可能是数十万个中的一个

的实体。

TempQuestions (Jia 等人, 2018a)是一个 KGQA 专门针对时间 QA 的数据集。它由来自 WebQuestions 的问题子集组成,Free917 (Cai 和 Yates, 2013)和复杂问题 (Bao 等人, 2016)是时间性的

推理	示例模板	示例问题
简单时间 {head}	什么时候占据了 {tail} 的位置	奥巴马什么时候担任美国总统
简单实体 {head}	在 {time} 获得了哪个奖项	布拉德·皮特 2001 年获得哪个奖项
之前/之后谁是{tail}	{type} {head}	奥巴马之前的美国总统是谁
第一个/最后一个	{head} 什么时候玩他们的 {adj} 游戏	梅西第一次参加比赛是什么时候
加入时间	谁在{event}期间担任过{tail}的职位 谁在二战期间担任过美国总统	

表 2:不同类型时间推理的示例问题。{head}、{tail}和{time}对应形式为事实的实体/时间戳（头、关系、尾部、时间戳）。{event}对应event中的实体事实例如。第二次世界大战。{type} 可以是 before/after 之一,而 {adj} 可以是 first/last 之一。请参阅第3.2节细节。

自然。他们给出了“时间问题”的定义并使用了某些触发词（例如

之前，之后）以及其他约束从这些数据集中过滤掉属于该定义的问题。然而,这个数据集包含只有 1271 个问题 仅用于评估以及它所基于的知识图谱（KG 的子集）FreeBase (Bollacker et al., 2008))不是一个临时的公斤。另一个缺点是 FreeBase 没有自2015年以来一直在积极开发中,因此其中存储的一些信息已经过时并且这是不准确的潜在来源。

2.2 时间 QA 算法

据我们所知,最近的 KGQA 算法 (Miller et al. 2016; Sun et al. 2019; Cohen et al. 2020; Sun et al. 2020)使用非时间 KG,即包含事实的 KG

形式（主语、关系、宾语）。将这些扩展到包含以下形式事实的时间 KG（主题、关系、对象、开始时间、结束时间)是一个不平凡的任务。TEQUILA (Jia 等人，2018b)是一种方法专门针对时间 KGQA。龙舌兰酒将问题分解并重写为非时间子问题和时间约束。

然后使用以下方法检索子问题的答案任何 KGQA 发动机。最后,TEQUILA 使用时间间隔上的约束推理来计算完整问题的最终答案。这种方法的一个主要缺点是使用预先指定的分解的模板,以及对实体具有时间约束的假设。

此外,由于它是为非时间 KG 设计的,因此不是将其应用于时间 KG 的直接方法事实是有时间范围的。

3 CRONQUESTIONS:新的时间 KGQA数据集

CRONQUESTIONS,我们的时态 KGQA 数据集由两部分组成:带有时间注释的知识图谱和一组自然语言问题

需要时间推理。

3.1 颞叶KG

为了准备我们的时间 KG,我们首先采取所有带有来自 Lacroix 等人提出的Wiki 数据子集的时间注释的事实。（2020）。我们删除了谓词“member”的一些实例运动队”以平衡 KG 因为这个谓词占了 50% 以上事实。时间戳被离散化为年。这导致了一个包含 323k 事实、125k 实体的 KG 和203关系。

然而,这种对事实的过滤错过了重要的世界事件。例如,KG子集

使用上述技术创建的内容包含实体第二次世界大战,但没有相关事实它告诉我们第二次世界大战何时开始或结束。

需要这些知识来回答诸如此类的问题”谁是世界上的美国总统第二次世界大战?”为了克服这个缺点,我们首先从 WikiData 中提取具有“start”的实体时间”和“结束时间”注释。从这组作品来看,然后我们删除了游戏节目的实体,电影或电视剧（因为这些不是重要的世界事件,但确实有开始和结束时间注释）,然后删除实体少于 50 个相关事实。这最后一套实体然后将事实添加为格式中的事实（WWII,重大事件,发生于 1939 年、1945 年）。决赛时间 KG 由 328k 个事实组成,其中 5k 是事件事实。

3.2 时间问题

为了生成 QA 数据集,我们从一组开始用于时间推理的模板。这些曾经是使用五个最常见的关系我们的 WikiData 子集,即

- 运动队成员
- 持有的立场
- 获得的奖项
- 配偶

模板	{head} 何时在 {tail} 中播放
种子Qn	梅西什么时候在巴塞罗那踢球
人类	梅西什么时候在巴塞罗那踢球
释义	梅西效力巴塞罗那足球俱乐部是哪几年的事 巴塞罗那队什么时候有梅西的 梅西在巴塞罗那足球俱乐部踢球的时间是什么时候
机器	梅西什么时候为巴塞罗那足球俱乐部效力
释义	梅西什么时候在巴塞罗那踢球 梅西什么时候在巴塞罗那踢过球

表 3:人类生成的空位填充释义和机器。详细内容请参见3.2节。

	火车开发	测试
简单实体	90,651 7,745 7,812	
简单的时光	61,471 5,197 5,046	
之前/之后	23,869 1,982 2,151	
第一个/最后一个	118,556 11,198 11,159	
时间加入	55,453 3,878 3,832	
实体答复	225,672 19,362 19,524	
时间答复	124,328 10,638 10,476	
合计	350,000 30,000 30,000	

表 4:我们的数据集中涉及不同推理类型和不同答案的问题数量类型。详细内容请参见3.2.1节。

· 雇主

这产生了超过 30 个独特的种子模板

五种关系和五种不同的推理结构（部分示例请参见表 2）。每一个

这些模板有相应的过程

可以在时间 KG 上执行以提取

该模板的所有可能答案。然而，

与张等人类似。（2017），我们选择不

使此过程成为数据集的一部分，以删除

质量保证系统对此类的不受欢迎的依赖

正式的候选人收集方法。这也是

允许轻松增加数据集，因为只有

需要问答对。

本着与 ComplexWebQuestions 相同的精神，

然后我们要求人类注释者解释

这些模板是为了生成更多的语言

多样性。为注释者提供了带有虚拟实体和时间的槽填充模板，并询问

重新表述问题，使虚拟实体/时间出现在释义中，并且

问题的意思没有改变。这导致了

246 个独特的模板。

然后我们使用Hu 等人开发的单语释义器。（2019）使用这 246

个模板自动生成释义。后

通过注释器验证它们的正确性，我们

最终得到 654 个模板。这些模板是

然后使用 WikiData 中的实体别名填充

生成 410k 个独特的问答对。

最后，将数据分为训练/测试

折叠，我们确保

1.火车问题的释义不存在

测试题。

2.试题之间不存在实体重叠

和训练问题。允许事件重叠。

第二个要求意味着，如果存在“奥巴马之前谁是总统”的问题

在训练集中，测试集不能包含任何提及实体“奥巴马”的问题。虽然这

策略可能看起来过于谨慎，它确保模型进行时间推理，而不是根据训练期间看到的实体进行猜测。刘易斯等人。（2020）注意到一个问题

他们发现近 30% 的 WebQuestions

测试问题与训练问题重叠。该问题已在 MetaQA 中发现

数据集也是如此，其中存在显着重叠

测试/训练实体和测试/训练问题之间

释义，即使使用部分 KG，也能在基线方法上获得令人怀疑的高性能

数据（Saxena et al., 2020），这表明表面上表现良好的

模型不一定是

对 KG 进行所需的推理。

我们的数据创建协议的一个缺点是

该问题/答案对是自动生成的。因此，从语义角度来看，问题分布是

人为的。（复杂的WebQuestions 也有类似的限制。）但是，

因为开发能够进行时间推理的模型是自然科学的一个重要方向

语言理解，我们认为我们的数据集

提供培训和评估的机会

KGQA 模型因其规模大而存在，尽管它具有低于自然的语言多样性。在

第6.4节，我们展示了训练数据的效果

尺寸对模型性能有影响。

总而言之，我们的每个示例都包含

1. 一个转述的自然语言问题。

2. 问题中的一组实体/时间。

3. 一组“黄金”答案（实体或时间）。

这些实体被指定为 WikiData ID（例如，Q219237），时间是年份（例如，1991）。我们

也包括测试问题中的实体/时间集，因为与其他 KGQA 数据集类似

（MetaQA、网络问题、复杂网络问题）

以及使用这些数据集的方法（PullNet、EmQL），实体链接被视为一个单独的问题，完整的实体链接如下：

总结道。我们还在训练折叠中包含种子模板和头/尾/时间注释,但在测试折叠中省略这些。

3.2.1 问题分类

为了便于分析,我们将问题分为“简单推理”和“复杂推理”问题(分布统计见表4)。

简单推理: 这些问题需要一个事实来回答,答案可以是一个实体或一个时间实例。例如“2008年美国总统是谁?”这个问题需要一个事实来回答这个问题,即(巴拉克·奥巴马,担任美国总统,2008年,2016年)

复杂推理: 这些问题需要多个事实来回答,并且可能更加多样化。

例如“谁是美国第一任总统?”这需要与“美国总统”实体相关的多个事实进行推理。在我们的数据集中,所有不是“简单推理”问题的问题都被视为复杂问题。

这些问题进一步分为“之前/之后”、“第一次/最后一次”和“时间加入”类型 - 请参阅表2了解这些问题的示例。

4 时间 KG 嵌入

我们研究如何使用时间和非时间的 KG 嵌入以及预训练的语言模型来执行时间 KGQA。我们将首先简要描述我们使用的具体 KG 嵌入模型,然后继续展示我们如何在 QA 模型中使用它们。在所有情况下,分数都会转化为不完整知识图谱中正元组和负元组的适当损失,并将这些损失最小化以训练实体、时间和关系表示。

4.1 复杂性

Complex (Trouillon et al., 2016) 将每个实体 e 表示为复向量 $u_e \in \mathbb{C}^D$ 。每个关系

r 也表示为复向量 $v_r \in \mathbb{C}^D$ 。所声称的事实 (s, r, o) 的得分 ϕ 为

$$\phi(s, r, o) = (\text{us}, v_r, u_o) = \sum_{d=1}^D \text{us}[d]v_r[d]u_o[d] \quad (1)$$

其中 (\cdot) 表示实部, c 是复共轭。尽管取得了进一步的发展, Complex 以及完善的培训协议

(Lacroix 等人, 2018) 仍然是最强的知识库嵌入方法之一 (Ruffinelli 等人, 2020)。

4.2 TComplex、TNTComplex

Lacroix 等人 (2020) 采取了早期措施来随着时间的推移扩展 Complex。每个时间戳 t 也表示为一个复向量 $u_t \in \mathbb{C}^D$ 。对于所声明的事实 (s, r, o, t) , 它们的 TComplex 评分函数为

$$\phi(s, r, o, t) = (\text{us}, v_r, u_o, u_t) \quad (2)$$

他们的 TNTComplex 评分函数使用关系 r 的两个代表 T 表示: 对时间敏感的 v_r 和对时间不敏感的 v_r 。评分函数是时间敏感部分和时间不敏感部分的总和: $(\text{us}, v_r, u_o, u_t)$

4.3 时间综合体

TimePlex (Jain 等人, 2020) 通过嵌入 $u_t \in \mathbb{C}^D$ 增强了 Complex, 用于离散时间点 t 。为了合并时间, TimePlex 对每个关系 r 使用三种表示形式, 即 $(v_r(s, r, o, t))$ 作为

所以 $(v_r, v_r^{ST}, v_r^{OT})$ 并写入元组的基本分数

$$\phi(s, r, o, t) = \text{us}, v_r^{ST} + \alpha \text{us}, v_r^{OT} + \beta \text{us}, v_r^{OT} + \gamma \text{us}, u_o, u_t \quad (3)$$

其中 α, β, γ 是超参数。

5 CRONKGQA: 我们提出的方法

我们从时间 KG 开始, 对其应用时间不可知或时间敏感的 KG 嵌入算法 (Complex、TComplex 或 TimePlex), 并获得时间 KG 的实体、关系和时间戳嵌入。我们将使用以下符号。
 \cdot E 是实体嵌入矩阵
 \cdot T 是时间戳嵌入矩阵
 \cdot ET 是 E 和 T 矩阵的串联。

这用于对答案进行评分, 因为答案可以是实体或时间戳。

如果实体/时间戳嵌入是 \mathbb{C}^D 中的复值向量, 我们将它们扩展为大小为 $2D$ 的实值向量, 其中前半部分是原始向量的实部, 后半部分是原始向量的复部分。

我们首先将 EmbedKGQA (Saxena et al., 2020) 直接应用于 Temporal KGQA 任务。在其最初的实现中, EmbedKGQA 使用 Complex (第4.1节) 嵌入, 只能处理非时间知识图谱和单一实体问题。

为了将其应用于 CRONQUESTIONS, 我们将问题中遇到的第一个实体设置为

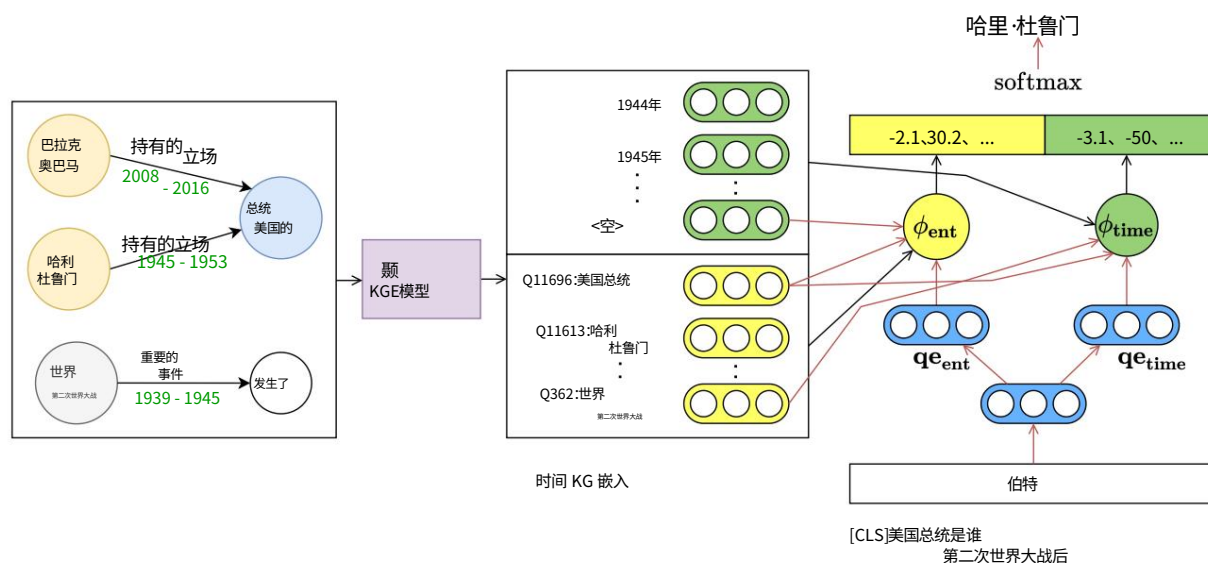


图 1: CRONKGQA 方法。(i) 时间 KG 嵌入模型 (第4 节)用于为时间知识图中的每个时间戳和实体生成 em 嵌入 (ii) BERT 用于得到两个问题

嵌入: q_{ent} 和 q_{time} 。(iii) 使用方程4和5将问题中提及的实体/时间嵌入与问题嵌入相结合,以获得实体和时间预测的得分向量。(iv) 得分向量为

使用 concatenated 和 softmax 来获取答案概率。详情请参阅第5节。

EmbedKGQA 需要“头实体”。随着

为此,我们将实体嵌入矩阵E设置为

Complex 嵌入我们的 KG 实体,并将T初始化为随机可学习矩阵。嵌入KGQA

然后对 ET 进行预测。

接下来,我们修改EmbedKGQA,使其可以使用时间 KG 嵌入。我们使用 TComplex

(第4.2 节)用于获取实体和时间戳 em 基础。CRONKGQA (图 1)利用两个

评分函数,一种用于预测实体,

一个用于预测时间。使用预先训练的 LM

(在我们的例子中为 BERT) CRONKGQA 发现一个问题

嵌入 q_e 。然后预计会得到两个

嵌入、 q_{ent} 和 q_{time} ,这是个问题

分别用于实体和时间预测的嵌入。

实体评分函数:我们从问题中提取主题实体s和时间戳t。如果

两者都缺失,我们使用虚拟实体/时间。

然后,使用评分函数 $\phi(s, r, o, t)$

等式2,我们计算每个实体的分数

和 ϵ 和

$$\phi_{ent}(e) = (us, q_{ent}, u \text{ 重量}) \quad (4)$$

其中E是KG中实体的集合。这

为我们每个作为答案的实体提供一个分数。

时间评分函数:同样,我们提取一个

问题中的主体实体s和客体实体o,如果不存在则使用虚拟实体。

然后,使用2,我们计算每次的分数

tamp $t \in T$ as

$$\phi_{time}(t) = (us, q_{time}, u \text{ 重量}) \quad (5)$$

所有实体和时间的分数被连接起来,并使用softmax来计算答案

该组合得分向量的概率。这

使用交叉熵损失来训练模型。

6 实验与诊断

在本节中,我们旨在回答以下问题

问题:

- 1.基线和CRONKGQA如何执行关于CRONQUESTIONS任务? (第6.2 节。)
- 2.某些方法是否比其他方法表现更好?具体的推理任务? (第6.3 节。)
- 3.训练数据集大小 (问题数量)对机器学习性能的影响有多大

模型? (第6.4 节。)

- 4.时间 KG 嵌入比非时间 KG 嵌入有什么优势吗? (第6.5 节。)

6.1 其他方法比较

Petroni 等人已经证明了这一点。(2019)和Rafel 等人。(2020)大型 LM,例如 BERT 及其变体,捕获现实世界的知识 (从他们的大规模百科全书式训练中收集

语料库),可以直接应用于诸如

作为质量检查。在这些基线中,我们没有专门将我们的时间 KG 版本输入到模型中

模型	Hits@1 问							点击率@10						
	全面的	问题类型 答案类型 复杂 简单 实体 时间						全面的	问题类型 答案类型 复杂 简单 实体 时间					
		0.052	0.077	0.06	0.086	0.05	0.082		0.048	0.086	0.05	0.082	0.048	0.086
伯特	0.071							0.213		0.205	0.225	0.192	0.253	
罗伯塔	0.07							0.202		0.192	0.215	0.186	0.231	
KnowBERT	0.07	0.083		0.051	0.081	0.048		0.201	0.189		0.217	0.185		0.23
T5-3B 0.081 EmbedKGQA		0.073		0.091	0.088	0.067	0.29	-	-	-	-	-	-	-
0.288 T-EaE-添加	0.278	0.286		0.057				0.672	0.632	0.725		0.85	0.341	
		0.257		0.306	0.313	0.213		0.663	0.614	0.729	0.662	0.665		
T-EaE-替换	0.288	0.257		0.329	0.318	0.231		0.678	0.623	0.753	0.668	0.698		
CRONKGQA 0.647		0.392		0.987	0.699	0.549		0.884	0.802	0.992	0.898	0.857		

表 5:基线我们的方法在CRONQUESTIONS数据集上的性能。中线以上的方法不要使用任何 KG 嵌入,而下面的则使用时间或非时间 KG 嵌入。Hits@10 不适用于 T5-3B,因为它是文本到文本模型并进行单个预测。请参阅详情见6.2节。

相反,我们期望模型具有真实世界

计算答案的知识。

BERT:我们用 BERT, RoBERTa 进行实验

(Liu 等人, 2019)和 KnowBERT (Peters 等人, 2019),这是 BERT 的一个变体,其中的信息来自知识库,例如 WikiData 和

WordNet 已被注入到 BERT 中。我们添加一个

预测头位于 [CLS] 标记之上

最后一层并对其进行softmax来预测

回答概率。

T5:为了应用T5 (Raffel et al., 2020)

对于时间 QA,我们将每个问题进行转换

在我们的数据集中采用“时间问题”的形式:

问题?。评估有两种情况:

1.时间回答:我们进行精确的字符串匹配

T5 输出和正确答案之间。

2.实体答案:我们比较系统输出

KG 中所有实体的别名。这

具有最小编辑别名的实体

到预测的距离 (Levenshtein, 1966)

文本输出被视为预测实体。

实体作为专家: Fevry 等人. (2020)提议

EaE,旨在整合实体的模型

将知识转化为基于 Transformer 的语言

模型。对于CRONQUESTIONS上的时间 KGQA,我们假设所

有接地实体和

Question1 中标记了提及的时间跨度。

我们将此模型称为T-EaE-add。我们尝试

EaE 的另一个变体, T-EaE-replace,其中

而不是添加实体/时间和 BERT 令牌

嵌入,我们替换 BERT 嵌入

实体/时间的实体/时间嵌入

提及2

1 这个假设可以通过使用 EaE 的早期模型来消除变压器阶段作为 NE 观察者和消歧者。

2附录A.1提供了我们的 EaE 实施的详细信息。

6.2 主要结果

表5显示了各种方法的结果

我们的数据集。我们看到基于大的方法

单独预训练的 LM (BERT, RoBERTa, T5),如

和 KnowBERT 一样,表现明显较差

与使用 KG嵌入 (时间或非时间)增强的方法相比。这大

概是

因为有特定于我们的 KG 嵌入

时间 KG 帮助模型专注于那些

实体/时间戳。在我们的实验中, BERT 的表现略好于

KnowBERT,尽管

KnowBERT 在其参数中包含实体知识。

T5-3B 在我们测试的 LM 中表现最好,

可能是因为参数太多

和预训练。

即使在使用 KG 嵌入的方法中,

CRONKGQA 在所有指标上都表现最好,

接下来是 T-EaE-replace。自从嵌入KGQA

具有非时间嵌入,其性能

答案是时间的问题非常低

与 BERT 相当 这是使用的 LM

我们的 EmbedKGQA 实施。

另一个值得注意的有趣的事情是

在简单推理问题上的表现。

CRONKGQA 远远优于简单的基线

问题,达到接近 0.99 的命中率@1,其中

T-EaE 的值要低得多 (0.329)。我们相信有

可能有以下几个原因:

1.使用 TComplEx 评分函数组合嵌入存在归纳偏差

CRONKGQA,与中使用的相同

创建实体和时间嵌入,从而

让简单的问题变得简单易懂

回答。然而,不依赖评分函数意味着 T-EaE 可以扩展到任何

KG 嵌入,而CRONKGQA 不能。

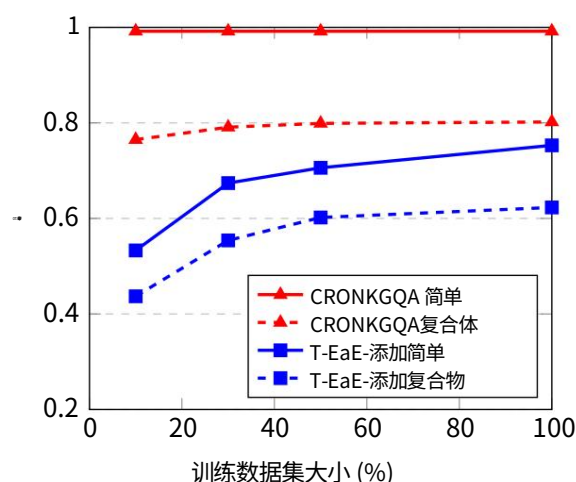


图 2:模型性能 (hits@10) 与训练 CRONKGQA 和 T-EaE 添加的数据集大小 (百分比)。实线用于简单推理,虚线用于简单推理线用于复杂推理类型的问题。为了每个数据集大小,模型都经过训练直至验证 10 个时期内, hits@10 没有增加。请参阅详细信息请参见 6.4 节。

2. 另一个促成因素可能是

需要训练的参数较少

CRONKGQA 而 6 层 Transformer 编码器需要在 T-EaE 中从头开始训练。

变压器通常需要大量

不同的数据才能成功训练。

6.3 跨问题类型的表现

表 6 展示了 KG 嵌入的性能

跨不同类型推理的基于模型。

如上文第 6.2 节所述, CRONKGQA 在简单推理问题上表现得非常好

(简单实体,简单时间)。在复杂问题类型中,所有模型 (EmbedKGQA 除外) 都表现出最佳的准时连接问题 (例如,“谁

与罗伯特·迪纳米特一起效力于巴西国家足球队”)。这是因为此类问题通常有多个答案 (例如所有

罗伯特·迪纳米特 (Roberto Dinamite) 比赛时的球员

对于巴西),这使得模型更容易

做出正确的预测。在其他两种问题类型中,答案始终是单个实体/时间。

之前/之后的问题似乎最具挑战性

所有方法,最好的方法只能实现

0.288 次点击@1。

6.4 训练数据集大小的影响

图 2 显示了训练数据集大小对模型性能。正如我们所看到的,对于 T-EaE-add,

将训练数据集大小从 10% 增加到

100% 稳步提高两者的性能

简单推理题和复杂推理题。这

CRONKGQA 中有些效果

复杂的推理,但简单的推理则不然

输入问题。我们假设这是因为

T-EaE 有更多可训练参数 它有

需要训练的 6 层变压器

刮擦 与需要的 CRONKGQA 相比

只是微调 BERT 并训练一些浅层的

投影层。这些结果证实了我们的假设,即拥有一个大型的 (即使是合成的) 数据集是

对于训练时间推理模型很有用。

6.5 时间 KG 与非时间 KG

嵌入

我们进行了进一步的实验来研究

时间与非时间 KG 嵌入的影响。我们替换了时间实体嵌入

在 T-EaE 中用 ComplEx 嵌入替换,并且

将时间戳视为常规标记 (不与任何实体/时间提及相关联)。

CRONKGQA CX 与 EmbedKGQA 相同。结果可以

如表 7 所示。正如我们所见,对于两者

CRONKGQA 和 T-EaE-replace,使用时间

与非时间 KGE (Complex) 相比, KGE (TComplex) 显著提高了性能。 CRONKGQA 获得了更大的提升

与 T-EaE-replace 相比在性能上,可能是因为评分函数已经建模

在 TComplex 之后而不是 ComplEx 之后,而有

T-EaE-replace 中没有这种特定于嵌入的工程。另一个观察结果是,具有时间答案的问题的得分非常低

两者的准确度 (分别为 0.057 和 0.062)

CRONKGQA-CX 和 T-EaE-replace-CX,其中

远低于这些模型所达到的效果

TComplex。这表明具有时间 KG

嵌入对于基于 KG 嵌入的方法获得良好的性能至关重要。

7 结论

在本文中,我们介绍了 CRONQUESTIONS,

用于时态知识图问答的新数据集。虽然存在一些时间

KGQA 数据集,它们都是基于非时间的

KG (例如 Freebase) 并且问题相对较少。我们的数据集包含时间 KG

以及大量需要各种推理结构的时间问题。为了

开发如此大的数据集,我们使用了合成

	前/后	第一的/最后的	时间加入	简单的实体	简单的的时间	全部
嵌入KGQA	0.199 0.324 0.223	T-EaE-添加	0.256	0.421	0.087 0.288	0.288
	0.285 0.175 0.296	0.321 0.278				
T-EaE-替换		0.256 0.288 0.168		0.318	0.346 0.288	0.288
CRONKGQA	0.288 0.371 0.511	0.988 0.985 0.647				

表 6:不同推理类型问题的 Hits@1。“简单实体”和“简单时间”对应简单表5中的问题类型,而其他对应于复杂问题类型。更多内容请参考6.3节细节。

问题类型	CRONKGQA	T-EaE-替换		
	CX TCX	CX TCX		
简单复杂	0.29 0.987 0.248	0.329		
实体答案	0.286 0.392 0.247	0.257		
	0.411 0.699 0.347	0.318		
时间答案	0.057 0.549 0.062	0.231		
全面的	0.288 0.647 0.247	0.288		

表 7: CRONKGQA 和 T-EaE-replace的 Hits@1 使用 ComplEx(CX) 和 TComplEx(TCX) KG 嵌入。更多详情请参阅第 6.5节。

生成过程,导致从语义角度来看是人为的问题分布。

然而,拥有大型数据集提供了训练模型的机会,而不仅仅是评估他们。我们通过实验表明,增加训练数据集大小稳步提高了某些方法在 TKGQA 任务上的性能。

我们首先应用基于大规模预训练 LM 的 QA 我们的新数据集上的方法。然后我们注入KG 时间和非时间的嵌入

这些 LM 并观察到显著的改进

在性能上。我们还提出了一种新方法,

CRONKGQA,能够利用时间

用于执行 TKGQA 的 KG 嵌入。在我们的实验中,

CRONKGQA 优于所有基线。

这些结果表明 KG 嵌入可以

尽管在复杂的推理问题上仍有很大的改进空间,但它可以有效地用于执行时间 KGQA。

致谢

我们要感谢匿名审稿人

感谢他们的建设性反馈,帕特·维尔加 (Pat Verga) 和

来自 Google Research 的 William Cohen 的富有洞察力的评论。我们还要感谢

Chitrang Gupta (印度理工学院孟买分校)在调试源代码和数据集方面提供的帮助。这部作品是

部分由谷歌研究中心的捐赠支持,印度和 Jagadish Bose 奖学金。

参考

Junwei Bao, Nan Duan, Zhao Yan, Ming Zhou, and 赵铁军. 2016. [基于约束的问题与知识图的转换](#).在诉讼程序中

COLING 2016,第26届国际会议
计算语言学:技术论文,
第 2503–2514 页,日本大阪。 2016年科林展
组织委员会。

库尔特·博拉克、科林·埃文斯、普拉文·帕里托什、蒂姆
斯特奇和杰米·泰勒。 2008. [Freebase:一个协作创建的图形数据库,用于构建人类知识](#)。 2008 年 ACM 会议论文集

SIGMOD 国际管理会议
数据,SIGMOD 08,第 1247–1250 页,纽约,
美国纽约。计算机协会。

安托万·博德斯、尼古拉斯·乌苏尼尔、苏米特·乔普拉和
杰森·韦斯顿。 2015. [大型简单题
用记忆网络回答](#)。 arXiv 预印本
arXiv:1506.02075。

安东尼·博德斯、尼古拉斯·乌苏尼尔、阿尔贝托·加西亚·杜兰、杰
森·韦斯顿和奥克萨娜·雅赫年科。
2013. [翻译嵌入以建模多关系数据](#)。在神经信息处理中

系统 (NIPS),第 1-9 页。

蔡青青和亚历山大·耶茨。 2013. [大型
通过模式匹配和词典进行语义解析
扩大](#)。计算语言学协会第 51 届年会记录 (第一卷:长论文),第 423-433
页,索菲亚,

保加利亚。计算语言学协会。

William W. Cohen, Haitian Sun, R. Alex Hofer 和
马修·西格勒。 2020. [可扩展的神经方法
使用符号知识库进行推理](#)。

Shib Sankar Dasgupta, Swayambhu Nath Ray 和
帕塔·塔鲁克达尔。 2018. [HyTE :基于超平面
时间感知知识图嵌入](#)。在
2018年实证会议论文集
自然语言处理方法,页面
2001-2011,比利时布鲁塞尔。协会
计算语言学。

雅各布·德夫林 (Jacob Devlin), 张明伟 (Ming-Wei Chang), 肯顿·李 (Kenton Lee) 和
克里斯蒂娜·图塔诺娃。 2019. [Bert :深度预训练
用于语言理解的双向转换器](#)。

Thibault Fevry,Livio Baldini Soares,Nicholas FitzGerald、Eunsol Choi 和 Tom Kwiatkowski。 2020.实体作为专家：实体的稀疏内存访问监督。 arXiv 预印本 arXiv:2004.07202。

阿尔贝托·加西亚-杜兰 (Alberto García-Durán)、塞巴斯蒂安·杜曼 (Sebastian Duman) 和马蒂亚斯·尼珀特 (Mathias Niepert)。 2018.学习序列编码器用于时间知识图的完成。 2018 年自然语言处理经验方法会议论文集,第 4816 页–

4821,布鲁塞尔,比利时。计算语言学协会。

里沙布·戈尔、赛义德·迈赫兰·卡泽米、马库斯·布鲁贝克、和帕斯卡·普帕特。 2019.历时嵌入用于时间知识图的完成。

J. Edward Hu,Huda Khayrallah,Ryan Culkin,Patrick Xia, Tongfei Chen, Matt Post, and Benjamin 范·杜尔姆。 2019.改进词法约束翻译解码和单语重写。 2019 年北方会议记录 计算语言学协会美国分会：人类语言技术，

第一卷（长论文），明尼苏达州明尼阿波利斯。计算语言学协会。

Prachi Jain,Sushant Rathi,Mausam 和 Soumen 查克拉巴蒂。 2020.时态知识库完成：新算法和评估协议。

2020 年实证会议论文集 自然语言处理方法（EMNLP），第 3733–3747 页，在线。计算语言学协会。

贾震,Abdalghani Abujabal,Rishiraj Saha Roy,Jan nik Strotgen 和 Gerhard Weikum。 2018a. 临时问题:临时问题的基准。在网络配套程序中

2018 年会议,WWW 18,第 1057–1062 页,Re public 和日内瓦州,CHE。国际的 万维网会议指导委员会。

贾震,Abdalghani Abujabal,Rishiraj Saha Roy,Jan nik Strotgen 和 Gerhard Weikum。 2018b. 龙舌兰酒。第 27 届 ACM 国际信息和知识管理会议论文集。

蒂莫西·拉克鲁瓦、威廉·奥博津斯基和尼古拉斯·乌苏尼尔。 2020.用于时态知识库补全的张量分解。 arXiv 预印本

arXiv:2004.04926。

蒂莫西·拉克鲁瓦、尼古拉斯·乌苏尼尔和纪尧姆·奥博津斯基。 2018.正则张量分解用于完成知识库。 arXiv 预印本 arXiv:1806.07297。

卡莱夫·利塔鲁 (Kalev Leetaru) 和菲利普·施罗德 (Philip A Schrodtt)。 2013.分享：有关事件、地点和语气的全球数据,1979 年– 2012年。 ISA 年会,第 2 卷,页数 1-49.西特西尔。

弗拉基米尔·我·莱文斯坦。 1966. 二进制代码能力纠正删除、插入和反转。在苏联物理文献,第 10 卷 (8),第 707 页– 710.苏联。

帕特里克·刘易斯、庞图斯·斯坦内托普和塞巴斯蒂安·里德尔。 2020.问答测试列车重叠开放域问答数据集。

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Man dar Joshi, Danqi Chen, Omer Levy, Mike Lewis, 卢克·泽特莫耶和维塞林·斯托亚诺夫。 2019. Roberta:一种稳健优化的 bert 预训练方法。

亚历山大·米勒、亚当·费什、杰西·道奇、阿米尔·侯赛因·卡里米、安托万·博德斯和杰森·韦斯顿。 2016.用于直接读取文档的键值存储网络。 arXiv 预印本 arXiv:1606.03126。

Qiang Ning, Hao Wu, Rujun Han, Nanyun Peng, Matt 加德纳和丹·罗斯。 2020.扭矩:读数时间排序问题的理解数据集。

马修·E·彼得斯、马克·纽曼、罗伯特·L·洛根 IV au2、罗伊·施瓦茨、维杜尔·乔希、萨米尔辛格和诺亚·史密斯。 2019.知识增强了上下文单词表示。

法比奥·彼得罗尼、蒂姆·罗克塔舍尔、帕特里克·刘易斯、安东·巴赫金、吴宇翔、亚历山大·H·米勒和塞巴斯蒂安·里德尔。 2019.语言模型作为知识库？

科林·拉斐尔、诺姆·沙泽尔、亚当·罗伯茨、凯瑟琳 Lee, Sharan Narang, Michael Matena, Yanqi Zhou, 李伟和彼得·J·刘。 2020.探索极限使用统一的文本到文本转换器进行迁移学习。

Pranav Rajpurkar,张健,Konstantin Lopyrev 和 珀西·梁。 2016.小队:100,000+ 个问题用于机器理解文本。 arXiv 预印本 arXiv:1606.05250。

丹尼尔·鲁菲内利、塞缪尔·布罗谢特和雷纳格穆拉。 2020.你可以教老狗新伎俩!关于训练知识图嵌入。在国际学习代表会议上。

Apoorv Saxena,Aditai Tripathi 和 Partha Taluk dar。 2020.改进多跳问答使用知识库 em床上用品进行知识图谱分析。在计算语言学协会第 58 届年会记录中，

第 4498–4507 页，在线。计算语言学协会。

海天太阳、安德鲁·O·阿诺德、塔尼亚·贝德拉克斯-韦斯、费尔南多·佩雷拉和威廉·W·科恩。 2020. 知识库查询的忠实嵌入。

海天孙、塔尼亚·贝德拉克斯-韦斯和威廉·W·科母鸡。 2019.Pullnet：通过知识库迭代检索进行开放域问答

文本。

阿隆·塔尔莫和乔纳森·贝兰特。2018。[网络作为回答复杂问题的知识库](#)。计算语言学协会北美分会2018年会议记录:人类语言技术,第1卷(长论文),第641-651页,路易斯安那州新奥尔良。计算语言学协会。

帕塔·普拉蒂姆·塔鲁克达尔、德里·维贾亚和汤姆·米切尔。2012. 关系事实的耦合时间范围。WSDM 2012 论文集。

西奥·特鲁永、约翰内斯·韦尔布尔、塞巴斯蒂安·里德尔、埃里克·高西尔和纪尧姆·布沙尔。2016.用于简单链接预测的[复杂嵌入](#)。在国际机器学习会议(ICML)中。

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, Nilesh Agrawal 和 Partha Talukdar。2020. In [teracte](#):通过增加特征交互来改进基于卷积的知识图嵌入。

AAAI 人工智能会议记录,第34卷(03),第3009-3016页。

Ashish Vaswani, Noam Shazeer, Nicky Parmar, Jacob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser 和 Illia Polosukhin。2017年,你需要的[就是关注](#)。

杨志林、齐鹏、张赛正、Yoshua Ben gao, William W Cohen, Ruslan Salakhutdinov 和 Christopher D Manning。2018. Hotpotqa :用于多样化、可解释的多跳问答的数据集。arXiv 预印本 arXiv:1809.09600。

叶文涛、张明伟、何晓东、高剑峰。2015. [通过分阶段查询图生成进行语义解析:使用知识库回答问题](#)。计算语言学协会第53届年会暨第七届自然语言处理国际联合会议论文集(第一卷:长论文),第1321-1331页,中国北京。计算语言学协会。

张宇宇、戴汉军、Zornitsa Kozareva, Alexander J. Smola 和 Le Song。2017. [知识图谱问答的变分推理](#)。

附录

A.1 实体作为专家 (EaE)

该模型架构遵循与实体内存层交错的Transformer ([Vaswani 等人, 2017](#))。它有两个嵌入矩阵,用于标记和实体。它适用于输入序列

x 如下。

$$\begin{aligned} X_0 &= \text{TokenEmbed}(x) \\ X_1 &= \text{Transformer}_0(X_0, \text{层数} = l_0) - \\ X_2 &= \text{实体内存}(X_1) \\ X_3 &= \text{层范数}(X_2 + X_1) \\ X_4 &= \text{Transformer}_1(X_3, \text{层数} = l_1) - \\ X_5 &= \text{任务特定头}(X_4) \end{aligned} \quad (6)$$

整个模型(变压器、令牌和实体嵌入以及特定于任务的头)使用实体链接、提及检测和掩码语言建模的损失进行端到端训练。

A.2 时间 KGQA 的 EaE

CRONQUESTIONS不提供用于训练语言模型的文本语料库,因此,我们将BERT ([Devlin 等人, 2019](#))用于Transformer0以及TokenEmbed (等式6)。

对于EntityMemory,我们使用实体和时间戳的 TComplex/TimePlex 嵌入,这些实体和时间戳已经使用CRONQUESTIONS KG进行了预训练(有关KG嵌入的详细信息,请参阅第4节)。修改后的模型如下: $X_1 = \text{BERT}(x)$

$$\begin{aligned} X_2 &= \text{实体时间嵌入}(X_1) \\ X_3 &= \text{层范数}(X_2 + X_1) \\ X_4 &= \text{Transformer}_1(X_3, \text{层数} = 6) - \\ X_5 &= \text{PredictionHead}(X_4) \end{aligned} \quad (7)$$

为了简单起见,我们假设所有接地实体和时间提及跨度都在问题中标记,即对于每个标记,我们都知道它属于哪个实体或时间戳(或者如果它不属于任何实体或时间戳)。因此,对于输入 x 中的每个标记 x_i , $X_1[i]$ 包含上下文 BERT 嵌入

$$x_i \cdot$$

对于 $X_2[i]$, 有3种情况。 – x_i 是对实体 e 的提及。那么 $X_2[i] = E[e]$ 。 – x_i 是对时间戳 t 的提及。那么 $X_2[i] = T[t]$ 。 – x_i 没有提及。那么 $X_2[i]$ 就是零向量。

PredictionHead将Transformer1的与BERT的[CLS]标记对应的标记的最终输出作为嵌入的预测答案。该答案嵌入使用点积针对ET进行评分,以获得每个可能答案的分数,并采用softmax来获得答案概率。该模型使用交叉熵损失在QA数据集上进行训练。我们会参考

将此模型称为T-EaE-add ,因为我们正在对BERT 和实体/时间嵌入进行逐元素求和。

T-EaE-replace我们没有添加实体/时间和 BERT 嵌入,而是用实体/时间提及的实体/时间嵌入替换 BERT em嵌入。具体来说,在等式第 4 步中输入Transformer1之前。 7, 1.如果xi不是实体或时间提及,则 $X3[i] = BERT(X1[i])$ 2.如果xi是实体或时间提及, $X3[i] = EntityTimeEmbedding(X1[i])$

模型的其余部分保持不变。

A.3 示例表8到12

包含来自CRONQUESTIONS验证集的一些示例问题,以及我们实验的模型的前 5 个预测。 T5-3B 具有单一预测,因为它是文本到文本模型。

问题	第二次世界大战前谁担任瑞典首相
问题类型之前/之后	
黄金答案	Per Albin Hansson
伯特	老埃米尔·斯坦格、西格德·易卜生、约翰·尼加兹沃尔德、莱拉·弗雷瓦尔兹、JS 伍兹沃斯
KnowBERT	贝尼托·墨索里尼、奥斯坦和恩、汉斯·迪特里希·根舍、温斯顿·丘吉尔、卢茨·施韦林·冯·克罗西克
T5-3B	把奶酪留在下面
EmbedKGQA	Per Albin Hansson、Tage Erlander、Carl Gustaf Ekman、Arvid Lindman、Hjalmar Branting
T-EaE-添加	佩尔·阿尔宾·汉森、曼努埃尔·罗哈斯、阿瑟·索维、康斯坦丁诺斯·德梅尔齐斯、卡尔·雷纳
T-EaE-替换	Per Albin Hansson、Tage Erlander、Arvid Lindman、Valere Bernard、Vladko Macek
CRONKGQA	Per Albin Hansson、Tage Erlander、Arvid Lindman、Carl Gustaf Ekman、Hjalmar Branting

表 8:推理类型问题之前/之后。

问题	《Man on Wire》何时获得奥斯卡最佳纪录片奖
问题类型简单时间	
黄金答案	2008
伯特	1995年、1993年、1999年、1991年、1987年
KnowBERT	1993, 1996, 1994, 2006, 1995
T5-3B	1997
嵌入KGQA	2017, 2008, 2016, 2013, 2004
T-EaE-添加	2008、2009、2005、1999、2007
T-EaE-替换	2009、2008、2005、2006、2007
克朗克质量保证	2008, 2007, 2009, 2002, 1945

表 9:带有时间答案的简单推理问题。

问题	约翰·艾伦·拉塞特在皮克斯工作期间与谁一起工作
问题类型加入时间	
黄金答案	Floyd Norman
伯特	蒂姆·库克、埃莉诺·温莎·利奇、大卫·R·威廉姆斯、罗伯特·M·博因顿、朱尔斯·斯蒂格
KnowBERT	1994、1997、华特迪士尼动画工作室、克里斯蒂安·库布里克、1989
T5-3B	约翰·艾伦·拉塞特
EmbedKGQA	约翰·拉塞特、弗洛伊德·诺曼、邓肯·马乔里班克斯、格伦·基恩、西奥多·泰
T-EaE-添加	约翰·拉塞特、安妮·玛丽·巴德韦尔、威尔·芬、弗洛伊德·诺曼、雷让·布尔达奇
T-EaE-取代	约翰·拉塞特、威尔·芬、弗洛伊德·诺曼、尼克·拉涅利、肯·邓肯
编年史	约翰·拉塞特、弗洛伊德·诺曼、邓肯·马乔里班克斯、大卫·普鲁克斯玛、西奥多·泰

表 10:时间连接类型问题。

问题	约翰·赫布利 (John Hubley) 在为工业电影公司工作之前在哪里工作
问题类型之前/之后	
黄金答案华特迪士尼工作室	
伯特	华特迪士尼工作室、华纳兄弟卡通公司、皮克斯、微软、美国海军
KnowBERT	理工学院、Piti e-Salpetri 医院、华特迪士尼影城、
T5-3B	伊丽莎白·布登布鲁克,耶鲁大学 伦敦电影学院
嵌入KGQA	华特迪士尼工作室、法兰西学院、华纳兄弟卡通公司、那不勒斯费德里科二世大学、苏黎世联邦理工学院
T-EaE-添加	华特迪士尼工作室、Fleischer Studios、UPA、Walter Lantz Productions、 韦尔斯利学院
T-EaE-替换	华特迪士尼工作室、纽约城市学院、UPA、 耶鲁大学、印第安纳大学
编年史	华特迪士尼工作室、UPA、圣彼得堡国立大学、 华纳兄弟漫画公司,法兰西学院

表 11:推理类型问题之前/之后。

问题	娜奥米·福纳·吉伦哈尔的最后一位结婚对象是
问题类型第一个/最后一个	
黄金答案 斯蒂芬·吉伦哈尔	
伯特	1928 年,詹妮弗·拉什、斯蒂芬·马洛里、马丁·兰道荣获巴伐利亚宪法金质奖章
KnowBERT	Nadia Benois、Eugenia Zukerman、德国国家足球队、Talulah Riley、Lola Landau
T5-3B	吉伦哈尔
嵌入KGQA	斯蒂芬·吉伦哈尔、娜奥米·方纳·吉伦哈尔、沃尔夫哈德·冯·博斯拉格、 海因里希·施威格、布鲁斯·帕特洛
T-EaE-添加	斯蒂芬·吉伦哈尔、玛丽安·佐夫、科特·史密斯、道格拉斯·怀尔德、格德·维斯佩曼
T-EaE-替换	斯蒂芬·吉伦哈尔、海蒂·布罗德莱特-亨克斯、娜奥米·福纳·吉伦哈尔、 小迈尔斯·科普兰 (Miles Copeland, Jr.),哥伦比亚众议院议员
编年史	斯蒂芬·吉伦哈尔、安东尼娅·弗雷泽、布鲁斯·帕特洛、 娜奥米·福纳·吉伦哈尔、沃尔夫哈德·冯·博斯拉格

表 12:第一个/最后一个推理类型问题。