

知识图上的复杂时态问答

Zhen Jia  
Southwest Jiaotong University, China  
zjia@swjtu.edu.cn

Rishiraj Saha Roy Max  
Planck 信息学研究所,德国 rishiraj@mpi-inf.mpg.de

Soumajit Pramanik IIT  
比莱,印度  
soumajit@iitbhillai.ac.in

Gerhard Weikum马克  
斯普朗克信息学研究所,德国 weikum@mpi-inf.mpg.de

抽象的

知识图谱问答 (KG-QA)是 IR 中的一个重要主题。具有时间意图的问题是一类具有实际重要性的特殊问题,但在研究中并未受到太多关注。这项工作提出了 Exaqt,这是第一个用于回答具有多个实体和谓词以及相关时间条件的复杂时间问题的端到端系统。 Exaqt 分两个阶段回答有关知识图谱的自然语言问题,一个阶段致力于高召回率,另一个阶段致力于顶级准确率。第一步计算 KG 中与问题相关的紧凑子图,并使用Group Steiner Trees 和微调的 BERT 模型,通过相关的时间事实明智地增强它们。第二步根据第一步的输出构建关系图卷积网络 (R-GCN) ,并通过时间感知实体嵌入和对时间关系的关注来增强 R-GCN。我们在 TimeQuestions 上评估Exaqt, TimeQuestions 是我们根据各种通用 KG-QA 基准编译的包含16 个时间问题的大型数据集。结果表明,Exaqt在回答 KG 上的复杂问题方面优于三个最先进的系统,从而证明了时间 QA 的专门处理是合理的。

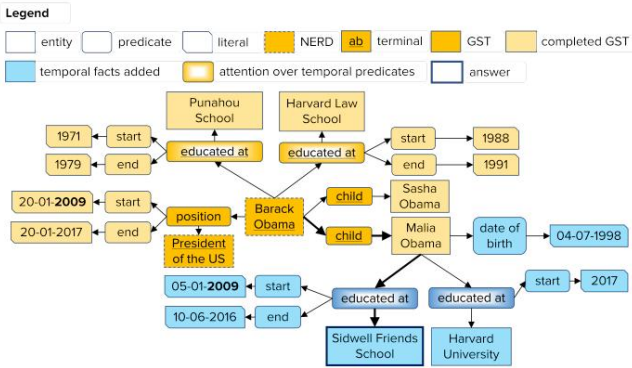


图 1:维基数据摘录,显示了奥巴马就任总统后他的孩子在哪里学习这一问题相关 KG 区域?答案是西德威尔友谊学校。

YAGO [64]或 DBpedia [10]已成为搜索引擎的首选资源,利用实体的结构化事实。虽然知识图谱问答[1,12,13,16,26,55,72,77,79 ]一直是一个主要话题,但很少有人关注临时问题的情况。此类问题涉及通过知识图谱中的关联时间戳来约束答案的显式或隐式概念。这涵盖了一个范围,从更简单的案例开始,例如奥巴马什么时候出生?、奥巴马 2001 年住在哪里?以及奥巴马在 9/11 期间住在哪里?更复杂的时间问题,例如:

奥巴马就任总统后,他的孩子们在哪里读书?

复杂问题必须考虑多跳约束 (Barack Obama → child → Malia Obama,Sasha Obama → 受教育于 → Sidwell Friends School) ,以及时间点和间隔交叉部分重叠的原因 (2009 年担任校长并在学校学习期间,2009年至2016年) 。图 1 显示了回答问题所需的 Wikidata KG 中相关区域的简化摘录。本文解决了复杂时间问题出现的这些挑战。

最先进的技术的局限性。关于非结构化文本源的时间 QA 的早期工作 [5,18,33,53,56,58,71 ]涉及各种形式的问题和文档解析,但没有延续到由实体和谓词组成的结构化事实的知识图谱。

专门针对 KG 的时间感知 QA 的少数作品包括[23,38,76 ]。 [38]使用一小组手工制定的规则进行问题分解和时间推理。这种方法需要人类专家来制定规则,并且不能处理复杂的问题。 [23]为捕获事件及其时间线的知识图谱构建了一个 QA 集合。 [76]中的键值记忆网络包含来自知识图谱的时间信息,用于回答简单的问题。

CCS 概念

信息系统→问答。

关键词

时态问答、复杂问题、知识图

ACM 参考格式: Zhen Jia, Soumajit Pramanik,Rishiraj Saha Roy 和 Gerhard Weikum. 2021.知识图上的复杂时间问答。第 30 届 ACM 国际信息和知识管理会议 (CIKM 21) 会议记录,2021 年 11 月 1 日至 5 日,虚拟活动,澳大利亚昆士兰州。 ACM,美国纽约州纽约市,11 页。 <https://doi.org/10.1145/>

nnnnnn.nnnnnnnn

1 简介

动机。具有时间信息需求的问题和查询[7,8,14,20,40 ]代表了搜索中的重要用例。对于事实问题,知识图谱 (KG)如维基数据[75],

允许免费制作部分或全部作品的数字或硬拷贝供个人或课堂使用,前提是制作或分发副本不是为了盈利或商业利益,并且副本在首页上附有此通知和完整引用。必须尊重本作品第三方组件的版权。

对于所有其他用途,请联系所有者/作者。  
CIKM 21,2021 年 11 月 1 日至 5 日,虚拟活动,澳大利亚昆士兰州 © 2021 版权所有/作者所有。  
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM。  
<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

方法。我们提出了 Exaqt:具有时间意图的复杂问题的可解释回答,这是一个不依赖手动规则来理解和推理问题的系统。 Exaqt分两步回答复杂的时间问题:(i)基于密集子图算法和微调的 BERT 模型,识别一个紧凑、易于处理的答案图,其中包含回答问题所需的所有线索;(ii)关系图卷积网络 (R-GCN) [66],用于推断图中的答案,并用有关时间的信号进行增强。

这两个阶段的工作原理如下(部分如图 1 所示)。

第一阶段:答案图构建。 Exaqt 获取问题中提到的实体的所有 KG 事实(美国总统巴拉克·奥巴马:虚线框),由现成的NERD 系统检测到[30,36,44]。通过微调的 BERT 模型(承认有关孩子 Malia 和 Sasha 的信息,但不承认 Michelle Obama 的信息),将由此产生的嘈杂事实集提炼成易于处理的事实集。为了构建所有问题相关 KG 项及其互连的KG 子图,计算组Steiner 树 (GST) [22, 47, 61] (深橙色节点、终端或下划线关键字匹配:“奥巴马”、“总统”、“孩子”、“受过教育”)并已完成(浅橙色节点)。此时的最后一步也是决定性的一步是用相关的时间事实来增强这个候选答案图,以引入有关相关日期、事件和时间相关谓词的线索(可能远离问题实体的多个跃点)。为此,我们使用类似的BERT 模型来识别与问题相关的时间事实(蓝色节点: Malia 和 Sasha 的教育背景及其日期)。

得到的答案图是第二阶段的输入。

第 2 阶段:R-GCN 的答案预测,受到流行的GRAFT-Net 模型[66]和相关工作 [59, 65] 的启发,我们构建了一个 R-GCN,它学习答案图上的实体嵌入,并将答案预测转化为节点分类任务。然而,先前作品中使用的 R-GCN 不了解时间约束[6]。为了克服这个障碍,我们通过时间感知实体嵌入、对时间关系的关注以及时间戳[80]、时间信号[60]和时间问题类别[38]的编码来增强R-GCN。在我们运行的示例中,时间注意力帮助 Exaqt 将注意力集中在与问题相关的关系(部分阴影节点)上。 Barack Obama的时间增强表示流经 R-GCN (粗边),并提高了Sidwell Friends School作为答案(粗边框节点)的可能性,其中包含 2009 年(粗体)的时间事实。通过为每个问题生成如此简洁的知识图谱片段(如图1所示),Exaqt 为其答案提供了可解释的证据。

贡献。这项工作做出了以下贡献:我们提出了Exaqt,第一个用于回答大规模知识图上的复杂时间问题的端到端系统。 Exaqt 应用经过微调的 BERT 模型和卷积图网络来解决识别复杂时间问题的相关 KG 事实的具体挑战;我们编制并发布了TimeQuestions,这是约16 个时间问题的基准(表1 中的示例);对完整Wikidata KG 的实验表明Exaqt 相对于三个最先进的复杂KG-QA 基线的优越性。该项目的所有资源均可在 <https://exaqt.mpi-inf.mpg.de/>和<https://github.com/zhenjia2017/EXAQT> 上获取。

类别问题	
显式的	谁赢得了 1986 年奥斯卡最佳女主角奖?
	2009 年雅科·范·多梅尔执导了哪部电影? 2012 德国使用什么货币?
隐含的	第九次十字军东征期间法国国王是谁?托马斯·杰斐逊在担任总统之前做了什么?克里斯蒂亚诺·罗纳尔多在曼联之后效力过哪个俱乐部?朱莉·安德鲁斯主演的第一部电影是什么?皮埃尔·德·顾拜旦担任的第二个职位是
	什么?伊丽莎白·泰勒最后一任丈夫是谁?
序数	
	湖人队在哪一年赢得了他们的第一个总冠军?
	温度。答。詹姆斯·卡格尼的配偶是什么时候出生的?
	金莺队上次赢得世界大赛是什么时候?

表 1:TimeQuestions 中的时间问题示例。

2 概念和符号

我们现在定义 Exaqt 背后的显着概念。

知识图谱。知识图(又名知识库)是组织为一组<主语、谓语、宾语>三元组的事实集合。它可以存储为此类三元组的 RDF 数据库,或者等效地存储为具有节点和边的图。例如Wikidata [75]、YAGO [64]、DBpedia [10]、Freebase [17] 和工业KG。当存储为图时,边是有向的:主语 → 谓语 → 宾语。主语和客体始终是节点,而谓词(又名关系)通常成为边缘标签。

事实。事实ε可以是二元的,包含通过谓词连接的主语和宾语,也可以是元的,通过主谓词和限定谓词组合多个项目。二元事实的一个例子是<Barack Obama, child, Malia Obama>,其中主语是实体(Barack Obama),对象可能是实体(Malia Obama),文字(常数,例如<Malia Obama, 出生日期>, 04-07-1998>),或类型又名类(<Sidwell Friends School中的私立学校, 类型,私立学校>)。我们在本文中可以互换使用术语谓词和关系。

一个事实结合了几个属于一起的三元组,例如< Barack Obama,担任的职位,美国总统;<Barack Obama,职位,美国总统;开始日期,2009 年 1 月 20 日;结束日期,2017 年 1 月 20 日>(见图 1)。持有的职位是主谓词,美国总统是主宾语,而其余数据是<限定词谓词,限定词宾语>对。事实在时间 QA 中至关重要,现代 KG 中的很大一部分时间信息被存储为限定符。在 KG 中表示限定符的一种方法如图 1 所示,通过从主谓词到限定符谓词再到限定符对象的路径。

时间事实。我们将时间事实ε定义为其中主要对象或任何限定符对象是时间戳的事实。例如<越南战争,结束日期, 1975 年 4 月 30 日>(二进制),或者,<史蒂文·斯皮尔伯格,获奖,奥斯卡最佳导演奖;工作的话,《辛德勒的名单》;时间点, 1993>(-ary)。

时态谓词。我们将时间谓词定义为可以将时间戳作为其直接对象或其限定符对象之一的谓词。例如出生日期和担任的职位。

时间问题。时间问题是包含时间表达式或时间信号,或其答案具有时间性质的问题[37]。时间表达的例子有“1998年”、“奥巴马总统任期”、“除夕夜”等,它们表明显式或隐式的时间范围[41]。时间信号[60]是

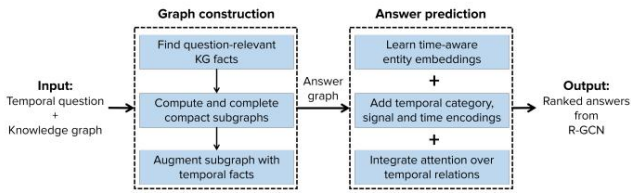


图 2: 两级 Exaqt 管道的概述。

时间关系的标记（之前、之后、重叠……）[6]，并用“之前、之后、期间……”等词语来表达，表明需要进行时间推理。在我们的模型中，问题表示为一组关键字  $\langle 1, 2, \dots \rangle$ 。

时间问题类别。时间问题分为四个基本类别[37]：(i) 包含明确的时间表达（“2009年”），(ii) 包含隐含的时间表达（“当奥巴马成为总统时”），(iii) 包含时间序数（“第一个”总统），以及 (iv) 有临时答案（“什么时候……”）。表 1 给出了几个时间问题的例子。一个问题可能属于多个类别。例如，朱莉·安德鲁斯与托尼·沃顿离婚后主演的第一部电影是什么？包含隐式时间表达式和时间序数。

回答。时间问题的答案是一组（可能是单例）实体或文字，例如，{Chicago University Lab School, Sidwell Friends School} 表示 Malia Obama 在哈佛之前在哪里学习？或 {08-2017} 表示 Malia 何时在哈佛就读？从哈佛开始？

答案图。答案图是知识图谱的子集，其中包含正确回答问题的所有必要事实。

3 构建答案图

图 2 是 Exaqt 的概述，有两个主要阶段：(i) 答案图构建（第 3 节）和 (ii) 答案预测（第 4 节）。

3.1 查找问题相关的知识图谱事实 NERD 问题实体。与大多数 QA 管道[16, 54] 一样，我们首先对输入问题运行命名实体识别和消歧 (NERD) [36, 44, 73]（奥巴马成为总统后，他的孩子在哪里学习？）。NERD 系统将问题中的单词范围识别为提及实体（“奥巴马”、“总统”），并将这些范围链接到 KG 项目或维基百科文章（可以轻松映射到流行的 KG）。这些相关实体（美国总统 Barack Obama）的事实为我们提供了知识库中的一个区域来开始寻找答案。

NERD 是 QA 轮中的关键齿轮：实体链接错误使主 QA 管道在答案检测方面无能为力。为了减轻这种影响，我们使用两个不同的系统 TagMe 和 ELQ [30, 44] 来提高答案召回率。复杂的问题通常包含多个实体提及，并且考虑到两个 NERD 系统，我们可以轻松地每个问题拥有 2-4 个不同的实体。因此，相关事实的总数可能是数百或更多。

为了将这一庞大且嘈杂的事实集减少为一些与问题相关的事实，我们对 BERT [24] 进行了如下微调。训练与问题相关的事实的分类器。对于训练集中的每个问题，我们运行 NERD 并检索检测到的实体的所有 KG 事实。然后，我们使用远程监督机制：在这些事实中，包含黄金答案的事实被标记为

作为正面实例。虽然几个复杂的问题可能在问题实体的事实中没有答案（多跳情况），但有答案的问题为我们的分类器提供了合理数量的问题相关性训练数据。请注意，还为问题实体检索带有限定词的事实（问题实体作为主语、宾语或限定词对象出现的完整事实）：这增加了我们获取正面示例的覆盖范围。

对于每个正例，我们从不包含答案的事实中随机抽取五个负例。对特定于问题的否定实例进行采样有助于学习更具辨别力的分类器，因为所有否定实例都保证至少包含问题中的一个实体（例如，<Barack Obama, 配偶, 迈克尔·奥巴马>）。使用所有不包含答案的事实将导致严重的类别不平衡，因为这远远高于正面实例的数量。

然后，我们将所有训练问题的<问题, 事实>配对的正面和负面实例汇集在一起。这对事实现在通过连接其组成部分被表达为自然语言句子；限定符语句使用“and”连接[50]。例如，奥巴马婚姻的完整事实（反例）是：<Barack Obama, 配偶, 迈克尔·奥巴马；开始日期, 1992 年 10 月 3 日；结婚地点, 基督教三一联合教会>。这两个限定词，可以表述为“Barack Obama 的配偶迈克尔·奥巴马，开始日期为 03-10-1992，结婚地点为基督教三一联合教会”。与口头事实配对的问题以及二进制真实标签被作为 BERT 序列对分类模型的训练输入。

应用分类器。按照[24]，问题和事实之间用特殊分隔符[SEP]连接起来，并在该序列前面添加特殊分类标记[CLS]。对应于[CLS]的最终隐藏向量，用  $\epsilon \in \mathbb{R}^d$ （是隐藏状态的大小）表示，被认为是累积表示。分类层的权重是微调期间引入的唯一参数，其中是类标签的数量（这里 = 2，事实是否与问题相关）。 $\log(\text{softmax}(\cdot))$  用作分类损失函数。一旦分类器经过训练，给定一个新的<问题, 事实>对，它就会输出与问题相关的事实概率（和标签）。我们对与某个问题相关的所有候选事实进行预测，并按该问题相关可能性的降序对它们进行  $\epsilon \in \mathbb{R}^d$ ，排序。我们从此处选择得分最高的事实  $\{ \}$  作为我们的问题相关集。

3.2 计算紧凑子图事实集  $\{ \}$  包含与问题相关的事实，但并不表明哪些是重要的连贯 KG 图集

对于这个问题，以及它们之间是如何联系的。为此，我们从上面的一组事实中归纳出一个如图 1 所示的图，其中每个 KG 项（实体、谓词、类型、文字）都成为自己的一个节点。边沿 KG 中规定的方向在同一事实的组件之间运行：主事实的主语  $\rightarrow$  谓词  $\rightarrow$  对象，以及（可选）限定符的主语  $\rightarrow$  谓词  $\rightarrow$  限定词谓词  $\rightarrow$  限定词对象。

注入连接。BERT 从我们的 NERD 系统检测到的许多实体的事实中选择  $\{ \}$ 。这些实体可能不会通过共享 KG 事实相互连接。然而，需要一个连通图，以便我们后续的 GST 和 R-GCN

算法可以产生预期的效果。为了在从 BERT 事实导出的图中注入连接性,我们计算每对问题实体之间的最短 KG 路径,并将这些路径添加到我们的图中。如果两个实体之间存在多个相同长度的路径,则按如下方式对它们的问题相关性进行评分。KG 路径是一组事实:长度为 1 的路径由一个事实组成 (Barack Obama → 职位 → 美国总统),长度为 2 的路径由两个事实组成 (Barack Obama ) → 国家 → 美利坚合众国 → 国家元首担任的职务 → 美国总统),等等。每个候选路径都被表述为一组事实 (分隔两个事实的句点)并用 BERT [39] 进行编码,问题也是如此。这些 BERT 编码存储在相应的[CLS]令牌中。我们计算[CLS] (问题)与[CLS] (路径)的余弦相似度,并将余弦相似度最高的路径添加到我们的答案图中。

商品及服务税模型。计算组斯坦纳树 (GST) [47,52,61,67] 已被证明是识别较大图中查询特定骨干结构的有效机制,例如,在数据库图上的关键词搜索中[4, 27],给定图中的节点子集 (称为终端),斯坦纳树 (ST) 是连接所有终端的成本最低的树。这减少到最低限度

当图的所有节点都是终端时生成树问题,以及当只有两个终端时的最短路径问题。GST 模拟了一种更复杂的情况,其中终端被排列成组或集合,并且足以找到连接每个组中至少一个节点的斯坦纳树。这个场景非常符合我们的需求,每个问题关键词都可以匹配图中的多个节点,自然地归纳出一个终端组。

找到一棵贯穿每个匹配节点的树是不现实的,因此需要组模型。

边缘成本。GST 问题的一个组成部分是如何定义边缘成本。由于边缘源自 KG 事实,因此我们利用问题) ∈ 由 Sec 的分类器分配的相关性分数。3.1: [0, 转换为边成本 1 - ) ∈ [0, 1]。 ( 1],

商品及服务税算法。GST 有很好的近似算法[45, 67],但 QA 需要高精度。因此,我们采用了Ding等人的固定参数易处理的精确算法。 [27]。它在较大的图上迭代地生长和合并较小的树,以达到最小的树。鉴于可能与问题无关的虚假连接,仅采用最好的树可能存在风险。

因此,我们使用了[27]的动态规划算法自然支持的顶部变体。

消费税完成。如图 1 所示,GST 生成连接最相关问题节点的骨架。为了将其转化为问题的连贯上下文,我们需要用构建该框架的事实来完成它。由于此步骤而引入的节点在图中以浅橙色显示:有关总统任期、奥巴马的孩子的日期以及有关奥巴马教育的 (嘈杂)事实。如果图具有多个连接的组件 (仍然有可能,因为我们之前的连接插入仅在实体上成对起作用),则为每个组件计算顶部 GST,并且联合图用于此事实完成步骤。

例子。我们在图 1 中展示了一个简化的示例,其中节点Barack Obama匹配问题关键字“Obama”, child匹配“children”, educated at匹配“study”, President of the United States匹配“president”。节点受教育者\_

与玛丽亚和萨莎有关的内容不在此出现,因为它们未包含在巴拉克·奥巴马的事实中,并且尚未出现在我们的答案图中。我们考虑精确匹配,不仅在节点标签中,而且在知识图谱中列出实体、谓词和类型的常见同义词的别名集中。这有助于我们考虑宽松的匹配,而不依赖于像word2vec [48]或 GloVe [51] 这样的模型,这些模型需要对相似性值进行不方便的阈值处理作为同义词的嘈杂代理。GST使用深橙色节点显示,并带有下划线的相关问题关键字匹配 (表示终端节点)。在实验中,我们只考虑实体的 NERD 匹配作为终端,以及其他 KG 项的关键字与别名的匹配。GST 自然包括连接终端所需的内部节点和边。请注意,出于 GST 计算的目的,该图被视为无向 (相当于双向)。

3.3 用时间事实增强子图获得所需答案图的最后一步是用时间事实增强它。在这里,我们在完整的 GST 中添加实体的与问题相关的时间事实。这会提取回答问题所需的时间信息,这些问题需要距离问题实体一跳以上的证据 (图 1 中的蓝色节点): <Malia Obama,在 Sidwell Friends School 接受教育;开始日期,05-01-2009> (+ 像 Malia 的出生日期这样的噪音)。此步骤背后的基本原理是捕获忠实回答问题所需的事实,其中忠实是指不是偶然而是在满足问题中的所有必要约束之后得出答案。

例如, 2016年莱昂纳多·迪卡普里奥获得哪届奥斯卡奖的问题?可以不用时间推理来回答,因为他只赢得了一次奥斯卡奖。我们希望在忠实的答复中避免此类情况的发生。

为此,我们首先从 KG 中检索已完成的 GST 中每个实体的所有时间事实。然后,我们使用类似的微调 BERT 模型来确定时间事实的问题相关性。该模型针对每个时间事实预测其包含答案的可能性。它使用包含答案的问题实体的时间事实作为正例进行训练,而负例是从这些时间事实中随机选择的。为了捕获网络中的多跳时间问题,我们探索问题实体的 2 跳事实以获得基本事实答案。在第一次微调期间没有使用更大的邻域,因为问题实体的两跳中的事实总数相当大,但两跳时间事实的计数是一个更容易处理的数字。此外,这符合我们对复杂时间问题的关注。令完整GST中实体的时间事实的似然得分为( )。和以前一样,我们取得分最高的{ },将它们添加到答案图中,然后传递到第 2 阶段。

#### 4 使用 R-GCN 预测答案

R-GCN 基础知识。Exaqt 的答案预测方法受到关系图卷积网络模型[59] 的启发,该模型是 GCN [29]的扩展,专门用于处理知识图等大规模关系数据。通常,GCN 对属于局部邻域的节点的特征 (相当于表示或嵌入向量)进行卷积,并将它们传播到最近的邻居。学习到的实体表示用于节点分类。这里,分类决策是节点是否是输入问题的答案。

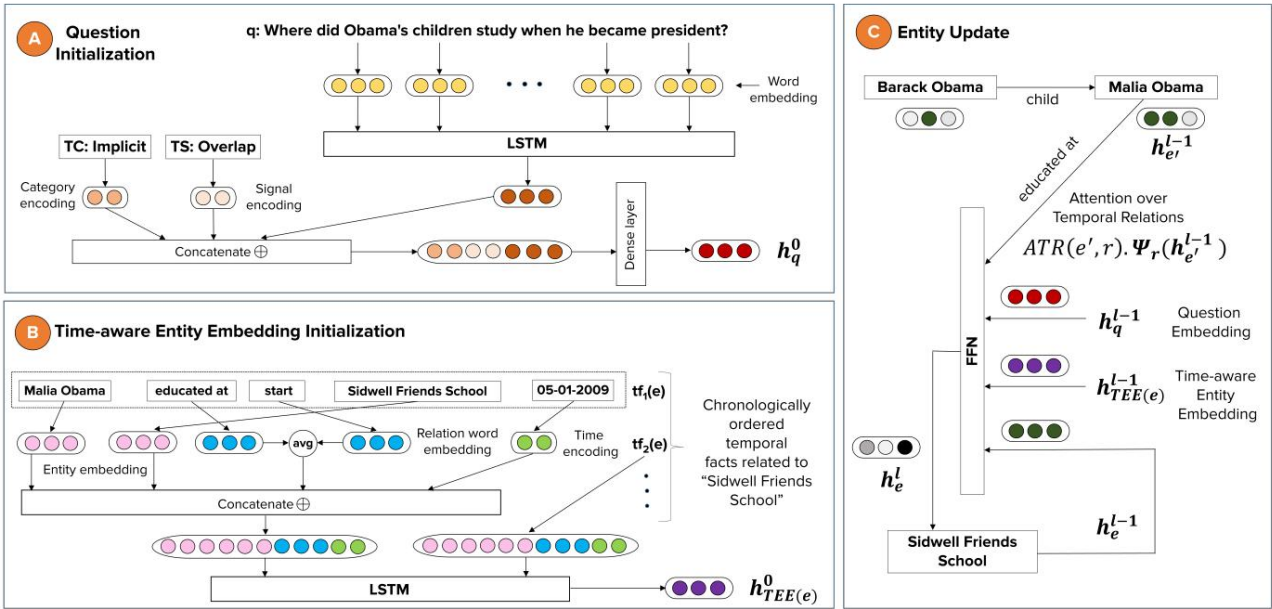


图 3:Exaqt 中的 R-GCN 模型架构,其中包括多个时间信息信号。

在这项工作中,我们使用广泛流行的 GRAFT-Net 模型[66],该模型采用 R-GCN 来处理 KG 和文本上的异构 QA [15, 50]。为了在我们的设置中应用这种答案预测机制,我们将上一步的答案图转换为有向关系图,并基于 GRAFT-Net 的唯一设置。在关系图中,实体、文字和类型成为节点,而谓词(关系)成为边标签。具体来说,我们通过具体化使用包含正常 SPO 三元组的 KG RDF 转储来获取二进制事实[35]。然后具体化的三元组可以直接表示为有向关系图[66]。Exaqt 在 GRAFT-Net 中的 R-GCN 上引入了四个主要扩展来处理时间 QA 任务: 我们嵌入时间事实来丰富实体节点的表示,创建时间感知实体嵌入(TEE); 我们对时间问题类别(TC)和时间信号(TS)进行编码以丰富问题表示; 我们采用时间编码(TE)来获得向量表示

多个类别)向量,其中每一位表示问题是否属于该类别。我们的标记器的工作原理如下: 如果标记器 SUTime [21]或 HeidelTime [62]检测到其中存在显式时间表达式,则问题将被标记为“EXPLICIT”类别。如果一个问题包含任何时间信号词(我们使用[60]编译的词典),并且满足某些词性模式,则该问题被标记为“隐式”类别; 如果问题以“何时...”、“哪一年...”和“在哪一天...”等短语开头,则该问题属于“临时回答”类型; 如果问题包含斯坦福CoreNLP 系统[9]标记的序数标签,以及某些关键字和词性模式,则该问题将被标记为“ORDINAL”类别。

时间戳的选项; 我们建议关注时间关系(ATR)来区分相同的关系但具有不同的时间戳作为对象。 接下来,我们将描述如何在扩展的R-GCN 架构中编码和更新节点表示并执行答案预测来处理时态问题。我们的神经架构如图 3 所示,而表 2 总结了此阶段使用的重要概念的符号。

4.1 问题表示4.1.1 初始化。为了对时间问题进行编码,我们首先确定其时间类别并提取时间信号(第 2 节)。

时间类别编码(TCE)。我们采用一种嘈杂但有效的策略来标记时间问题的类别,并将更复杂的(多标签)分类留作未来的工作。

我们使用四位多热点(回想一下,一个问题可以属于

时间信号编码(TSE)。Allen 的时间推理区间代数中定义了13 种时间关系,即: “等于”、“之前”、“相遇”、“重叠”、“期间”、“开始”和“完成”,其中除了“等于”之外,它们各自的倒数。

我们简化这些关系并将[37]中的策略调整为7大类时间信号: “之前”和“相遇”关系被视为“之前”信号; “前-逆”和“相遇-逆”关系被折叠为“后”信号; “开始”和“结束”关系分别映射到“START”和“FINISH”信号; 带有序数标签和“last”的单词映射到“ORDINAL”; 所有其他关系均被视为“重叠”信号; 缺少任何信号字会触发“无信号”情况。

我们使用字典将信号词映射到问题中的时间信号。然后,我们使用 7 位(一个问题可以包含多个信号)向量对这些信号进行编码,其中每位表示特定时间信号的存在或不存在。

除了这些时间类别和时间信号之外,我们还使用长短期记忆网络(LSTM)来建模



问题中的单词作为一个序列（参见图 3 中的块 A）。全面的，我们用  $\mathbf{v}$  来表示一个问题  $q$  的话如：

$$\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

这里  $\mathbf{v}_i$  和  $\mathbf{v}_j$  是编码时间类别和时间信号的多热向量，它们存在并表示

的预训练词嵌入（来自 Wikipedia2Vec [78]）。词在  $\mathbf{v}_i$  和  $\mathbf{v}_j$  向量。我们连接  $\mathbf{v}_i$  和  $\mathbf{v}_j$  向量。LSTM 最终状态的输出向量。最后，我们通过前馈网络传递这个串联向量 (FFN) 并获得初始嵌入表示为  $\mathbf{h}_0$ 。

4.1.2 更新。在后续层中，问题的嵌入使用属于它的实体的嵌入进行更新（即从NERD获得的问题实体）如下：

$$\mathbf{h}_i = \left( \sum_{e \in \mathcal{E}_i} \mathbf{h}_e \right)^{-1} \quad (2)$$

在哪里  $\mathcal{E}_i$  包含问题的实体并记录第  $i$  层实体的嵌入。 $\mathbf{h}_e^{-1}$  的

4.2 实体表示

4.2.1 初始化。用于初始化关系中的每个实体

图，我们使用固定大小的预训练嵌入，也来自 维基百科2Vec [78]。除了传统的 Skip-Gram 和上下文模型之外，Wikipedia2Vec 还利用了 Wikipedia 链接图：通过预测相邻实体来学习实体嵌入 维基百科图，产生更可靠的实体嵌入：

$$\mathbf{h}_0 = \dots \quad (3)$$

4.2.2 更新。在理解后续层实体的更新规则之前，我们需要介绍以下内容

概念：(i) 时间编码（TE）；(ii) 时间感知实体嵌入（TEE）；(iii) 对时间关系的关注（ATR）。时间编码（TE）。时间作为排序序列与文本中单词的位置具有固有的相似性：因此，我们采用正弦位置编码方法[74, 80]来表示时间戳

。在这里，位置（日、月等）将被编码为：

$$\mathbf{v}_t = \begin{cases} \sin(\frac{t}{10000^{\frac{2k}{D}}}), & \text{如果 } k = 2 \\ \cos(\frac{t}{10000^{\frac{2k}{D}}}), & \text{如果 } k = 2 + 1 \end{cases} \quad (4)$$

其中  $D$  是时间编码的维度， $k$  是（偶数/奇数）维向量中的位置。进一步，我们表示  $\mathbf{v}_t$ ，即时间编码作为编码的总和，每个对应的位置。这次的编码方式为每个时间戳提供唯一的编码并确保时间戳之间的顺序排序[80]，这对于推理至关重要 诸如时间问题中的“之前”和“之后”之类的信号。时间感知实体嵌入（TEE）。存在于的实体关系图与许多时间事实相关联  $\mathbf{h}_1, \mathbf{h}_2, \dots$ （第 2 节）在我们的答案图中。一个时间事实是如果出现在任何位置，则表示与某个实体相关联 事实的（主语、客体或限定词客体）。我们将每个编码为其实体嵌入、关系嵌入的串联 时间戳的（平均）和时间编码（如块所示）图 3 中的 B）。此外，我们按时间顺序排列  $\{\mathbf{h}_i\}$  中的每个事实对它们进行排序并通过 LSTM 网络传递它们。最后，输出

从 LSTM 的最终状态可以用作时间感知 的实体表示  $\mathbf{h}_i$ ，TEE  $\mathbf{t}_i$ ，这对于推理至关重要 R-GCN 模型：

$$\mathbf{h}_i = (\mathbf{h}_i^0, \mathbf{h}_i^1, \dots, \mathbf{h}_i^D) \quad (5)$$

在后续层中， $\mathbf{h}_i$  的嵌入将更新为其组成实体的嵌入得到更新。对时间关系的关注（ATR）。在时间 QA 中，我们需要区分与相同关系关联的实体，但是具有不同的时间戳（具有相同时间谓词的事实但不同的对象，就像一个人受过几个事实教育）。我们因此在这里引入时间注意力的概念，调整 GRAFT-Net [66] 中关系关注的更一般概念。在计算连接关系上的时间注意力时与实体，我们连接相应的关系嵌入 与其时间戳对象的时间编码并计算其 与该阶段问题嵌入的相似度：

$$\mathbf{a}_i = \left( \mathbf{h}_i \oplus \left( \mathbf{h}_i \right)^{-1} \right) \quad (6)$$

其中，softmax 归一化位于来自的所有传出边缘上， $\mathbf{a}_i$  是关系的预训练关系向量嵌入（Wikipedia2Vec 嵌入在 KG 的每个单词上 平均 predicate）， $\mathbf{t}_i$  是时间戳的时间编码 与 关系 相关联。对于与任何关系都没有关联的关系 时间戳，我们使用随机向量将其组合在一起。我们  $\mathbf{a}_i$ 。现在可以指定更新 涉及单层 FFN 的实体节点规则 以下四个状态的串联（参见图 3 的块 C）：

$$\mathbf{h}_i = \left( \mathbf{h}_i^{-1}, \mathbf{h}_i^{-1}, \mathbf{h}_i^{-1}, \mathbf{h}_i^{-1} \right) \quad (7)$$

这里，(i) 第一项对应于实体的表示 从上一层开始；(ii) 第二项表示前一层的问题表示；(iii) 第三项 表示前一层的时间感知  $\mathbf{en}(\mathbf{h}_i)$  的表示；(iv) 第四项合计 实体表示实体邻居的状态。在第四个任期内， 特定于关系的邻域对应于通过关系  $(\mathbf{r}_i)$  连接的实体集合，是对的关注

时间关系， $\mathbf{r}_i^{-1}$  是特定于关系的变换，具体取决于边的类型和方向：

$$(\mathbf{r}_i)^{-1} = \mathbf{r}_i^{-1} \cdot (\mathbf{r}_i, \mathbf{r}_i^{-1}) \quad (8)$$

这里  $\mathbf{r}_i^{-1}$  是个性化 PageRank [34] 分数，获得于 与 Graft-Net [66] 中相同的方式来控制传播 沿着从问题实体开始的路径嵌入。

4.3 答案预测

最终的实体表示（在二元分类设置）在层获得，则 中用于选择答案。对于每个 实体，我们将其概率定义为 的答案：

$$(\mathbf{e} \{ \} | \mathbf{h}_i, \mathbf{t}_i) = (\mathbf{h}_i + \mathbf{t}_i) \quad (9)$$

其中  $\mathbf{e} \{ \}$  是问题 的真实答案集，是为根据其答案图进行回答而构建 的关系图，

符号概念	
	层实体的表示 层问题的表示 问题的时间类别编码 问题的时间信号编码 从 NER 获得的问题实体
( )	
( )	
( )	
,	预训练的实体 ( ) 和关系 ( ) 嵌入
( )	时间戳的时间编码。 。 。按时间顺序
1, 2,	序排列的时间事实,用于表示 at 的时间事实,注意与 的关系特定变换相关的时间关系的时间感知实体表示
( )	
( , ) ( )	

实体的个性化 PageRank 分数位于

表 2:Exaqt 的 R-GCN 中概念的符号。					
类别	显式	隐式	温度	答	序数总计
免费Q [19]	44	4	76	11	135
网络Q [13]	315	77	112138	113	第788
复杂Q [11]	217	131	43	33	第424
图Q [63]	264	30	13	42	第349
复杂WebQ [68]	1356	224	595	315	2490
康品质量保证[2]	669	355	1180	1587	3791
LC-QuAD [69]	122	19	0	26	167
LC-QuAD 2.0 [28]	3534	第636	3726	819	8715
全部的	6521	第1476	5916	2946	16859

表 3:时间问题中按来源划分的问题类型分布。总和16859超出题数

16181 ,因为有些问题属于多个类别。

是 sigmoid 激活函数。和 分别是  
对应于分类器的权重和偏差向量是  
使用这些概率的二元交叉熵损失进行训练。

5 实验设置

5.1 基准测试

以前关于时间问题的集合,TempQuestions [37]  
Event-QA [23]每个只包含大约一千个问题,  
并且不适合构建神经模型。我们利用最近的  
社区在 QA 基准测试方面的努力,我们搜索  
用于与时间相关的问题的八个 KG-QA 数据集。结果是  
新汇编,TimeQuestions,有 16, 181 个问题,我们  
与本文一起发布 (详细信息见表3) 。由于其中一些  
之前的基准测试是通过 Freebase 或 DBpedia 进行的,我们使用  
这些知识图谱中的维基百科链接将它们映射到最大的维基数据  
当今发展最积极的公共 KG,也是我们使用的  
在这项工作中。每个基准测试中的问题都被标记为时间性的  
使用 SUTime [21]和 HeideTime [62] 的表达式,以及信号  
使用由[60]编写的词典的单词。每当有问题时  
发现至少有一个时间表达或信号词,它  
成为候选时间问题。作者对这个候选集 (大约 20 个问题)进行了误  
报过滤。对于每个  
对于这些问题,作者手动验证了正确性  
答案,如果不正确,则用正确的答案替换。此外,每个问题都手动标记  
了其时间问题  
可能的类别 (显式、隐式、时间答案或序数)  
帮助构建针对时间问题的自动分类器,

子问题本身就很有趣。我们分割我们的基准

以 60 : 20 : 20 的比例创建培训 (9708 个问题) ,  
开发 (3236)和测试 (3237)集。

5.2 基线

我们使用以下复杂 KG-QA 的最新方法作为基线来与 Exaqt 进行比较。所有  
基线均分别使用 TimeQuestions 的训练集和开发集进行训练和微调。

它们是最自然的基线选择,因为 Exaqt 受到启发

通过这些方法中的组件来构建其管道:  
Uniqorn [52]展示了 GST 在复杂 KG-QA 中的有效性,  
GRAFT-Net [66]和 PullNet [65]显示了 R-GCN 的价值  
用于答案预测。这些技术是为处理  
具有异构答案源 (知识图谱和文本) ,我们使用  
他们的仅 KG 变体:

- Uniqorn [52]:这是一种使用 Group Steiner Trees 回答复杂问题的方法,是 [47] 的扩展;
- GRAFT-Net [66]:这是第一个适应 R-GCN 的技术  
用于异构源的质量保证;
- PullNet [65]:该算法将 GRAFT-Net 分类器扩展到多跳问题的场景。由于  
代码不公开,我们使用了重新实现。

5.3 指标

所有系统都会返回答案的排名列表,其中包含与唯一标识符关联的  
KG 实体或文字。因此我们使用  
以下用于评估 Exaqt 和基线的指标,平均  
关于基准测试中的问题:

- P@1:如果排名最高,则最高排名的精度为 1  
答案正确,否则为零。
- MRR:这是第一等级的倒数,其中我们有  
正确答案。如果正确答案没有出现在  
排名列表,MRR 为零。
- Hit@5:如果第一个出现正确答案,则设置为 1  
五个位置,否则为零。

5.4 初始化

配置。我们使用 Wikidata KG 转储(https://dumps.  
wikimedia.org/wikidatawiki/entities/)从 4 月份开始采用 NTriples 格式  
2020 年,包含 12 个三元组,未压缩时容量为 2 TB  
在磁盘上。我们随后删除了语言标签、外部 ID、  
转储中的架构标签和 URL,给我们留下了大约 2 个三元组,消耗了 340 GB 磁  
盘空间。

对于 BERT 微调,正例和负例是根据 TimeQuestions 训练集和开发集  
以 1:5 的比例创建的。  
这些实例按 80 : 20 的比例组合和拆分 (测试  
不需要设置) ,其中第一个分割用于训练,而  
第二个分别用于超参数选择,用于 BERT 微调。我们使用 BERT-base-cased  
模型进行序列对分类(https://bit.ly/3fRVqAG)。微调的最佳参数

分别是:累积= 512,时期数= 2,丢失= 0.3,  
小批量大小= 50,权重衰减= 0.001。我们使用 AdamW 作为  
学习率为 $3 \times 10^{-5}$ 的优化器。作答期间图  
构建时,我们使用前 25 个与问题相关的事实( $|I| = 25$ ),  
前 25 个 GST (= 25) 和前 25 个时间事实 ( $|I| = 25$ )。

R-GCN模型训练。问题词、关系 (KG 谓词)词和实体的 100 维嵌入是从 Wikipedia2Vec [78]获得的,并从 2021 年 3 月的维基百科转储中学习。TCE、TSE、TE 和 TEE 的维度 (第 4 节)也全部设置为 100。使用了 LSTM 的最后隐藏状态

作为适用的编码。这是在 Nvidia Quadro RTX 8000 GPU 服务器上进行训练的。在 TimeQuestions 开发集上调整超参数值:GCN 层数 = 3、 epoch 数 = 100、小批量大小 = 25、梯度剪辑 = 1、学习率 = 0.001、LSTM dropout = 0.3、线性 dropout = 0.2,事实丢失 = 0.1。使用ReLU激活函数。

6 项重要发现

Exaqt 的回答性能和基线见表 4 (粗体列中的最佳值) 。主要观察如下。

Exaqt 优于基线。表 4 的主要观察结果是 Exaqt 相对于基线的全面优势。

每个类别、基线和度量的统计显着结果表明,通用的复杂 QA 系统无法处理时间 QA 的挑战性要求,并且需要时间增强的方法。超越每个基线可以提供单独的见解,如下所述。

商品及服务税 (GST) 还不够。 GST 是复杂 QA 的强大机制,可识别 KG 子集中的主干骨架,并从噪声图中删除不相关的信息。虽然这促使使用 GST 作为 Exaqt 中的构建块,但优于Uniqorn [52]方法表明 GST 中的非终结符 (内部节点)本身不足以回答时间问题。

用时间信息增强 R-GCN 效果很好。 R-GCN 是一个强大的模型,这一事实从以下事实可以清楚地看出: Graft-Net 在没有任何对时间 QA 的明确支持的情况下,成为这一具有挑战性的设置中最强的基线。我们工作的核心贡献是用不同类型的 R-GCN 来扩展

时间证据。对 GRAFT-Net 的改进表明,我们的多管齐下机制 (TEE、ATR、TCE、TSE 和 TE)成功地将R-GCN 模型的范围扩展到具有时间意图的问题。消融研究 (第 7 节)表明,每个“分支”在 Exaqt 的整体性能中都发挥着积极的作用。

并非每个问题都是多跳的。 PullNet 是一个最先进的系统,用于回答多跳链连接问题 (奥巴马的父亲在哪里出生? )。作为对 Graft-Net 的改进而提供的 PullNet 在我们的设置中存在不足,这可能看起来很奇怪。在检查示例中,原因很清楚 :PullNet 假设所有答案都位于问题实体的 2 跳圆周上 (理想情况下, -hop,其中 是需要为基准固定的变量:1 是过于简单化,而3 对于大 KG 来说是棘手的,因此我们选择 2 作为时间问题) 。如果情况并非如此 (例如,当答案位于 2 跳事实的限定词中时,会出现稍微棘手的情况:奥巴马的孩子何时开始在西德韦尔朋友学校学习?或者问题很简单:奥巴马何时出生? ) ,PullNet 无法利用这个训练点,因为它依赖于问题和答案实体之间的最短 KG 路径。这种统一跳跃假设并不总是实用,并且不能推广到PullNet 训练和评估之外的情况。

时间类别因难度而异。我们使用基准测试中问题类别的手动地面实况标签来进行深入研究

分类结果 (第 4.1.1 节中的噪声标记器具有  $\approx 90\%$  的准确度) 。具有临时答案的问题显然是最简单的。

请注意,这包括以“何时”开头的问题,许多模型使用专用词汇答案类型来处理这些问题[3, 12],类似于“哪里……?”的位置类型答案。问题。具有明确时间表达的问题是阶梯的下一级:虽然它们确实需要推理,但明确的年份通常会使这种匹配更容易 (谁在 1989 年成为南非总统? )。带有隐式表达式的问题更具挑战性:我们相信这才是 R-GCN 真正发挥作用的地方,因为基于 GST 的 Uniqorn显然存在不足。最后,时间序数的问题似乎超出了图神经网络中隐式推理的处理能力:当  $P@1 < 0.5$  时,它们构成了最大的研究挑战。

我们认为,这需要重新审视符号推理,最好将其插入神经 GCN 架构中。

7 深入分析

书呆子变种。我们尝试了 TagMe [30]、 AIDA [36]和 ELQ [44],从最流行的到最新的选择。各种选择的效果如表5所示。我们的最佳配置是TagMe + ELQ。 TagMe (在修剪实体上没有阈值的情况下使用)和 ELQ (使用默认参数运行)很好地互补,因为一个是面向召回的 (TagMe) ,另一个是偏向精度的 (ELQ) 。答案回忆衡量最终答案图 (测试集)中至少出现一个黄金答案的问题的比例。 AIDA + ELQ 检测到每个问题的实体数量相似,但在答案召回方面稍差一些。

理解阶段1。遍历Exaqt 面向召回的图构建阶段的步骤,我们尝试了解我们在哪里获得 (和失去)时间问题的答案 (表 6,测试集) 。首先,我们发现即使是两个 NERD 系统也无法保证完美的答案召回率 (75.8%) 。从第 1 行下降到第 2 行是预料之中的,因为人们无法在由第 1 行的所有事实引发的如此大的图上有效地计算图算法。添加最短路径 (第 3 行) ,同时使答案图更加连接 (之前: 1.58每个问题的连接组件,之后: 1.16) ,也对将正确答案带入图中略有帮助。从第 4 行和第 5 行,我们看到,从每个连接的组件中取顶部 (= 25)个 GST 的并集证明是值得的 (从 0.613 增加到0.640) ,完成 GST 也是如此 (进一步上升到0.671) 。

最后,添加时间事实提供了关键的提升,使第一阶段结束时的答案召回率达到了可观的72.4%。这意味着 2343 个问题的答案在图中被传递到 R-GCN (参见 GRAFT-Net 基于 PPR 的答案图中存在 1989 个答案) ,其中 1830 个最终得到正确答案。第二列计算答案图中实体和文字的平均数量 (候选答案) ,对于了解每个步骤的图形大小及其在答案召回方面的潜在权衡非常有洞察力。

理解第2 阶段。我们进行了消融研究,以了解Exaqt 的面向精度的第 2 阶段 :R-GCN 答案分类器中各个时间分量的相对影响。表 7 显示了测试集上的 P@1 结果,其中完整模型总体上以及每只猫都获得了最佳结果

埃戈里。完整模型 (第 1 行)的下降量表明特定组件的重要性程度。最重要的增强是对时间关系的关注 (ATR) 。全部



类别	全面的			显式的			隐含的			温度,年。			序数		
方法	P@1 MRR 命中@5 P@1 MRR 命中@5 P@1 MRR 命中@5 P@1 MRR 命中@5 P@1 MRR 命中@5														
独角兽[52]	0.331	0.409	0.538	0.318	0.406	0.536	0.316	0.415	0.545	0.392	0.472	0.597	0.202	0.236	0.356
移植网[66]	0.452	0.485	0.554	0.445	0.478	0.531	0.428	0.465	0.525	0.515	0.568	0.660	0.322	0.313	0.371
拉网[65]	0.105	0.136	0.186	0.022	0.043	0.075	0.081	0.123	0.192	0.234	0.277	0.349	0.029	0.049	0.083
确切地	0.565 * 0.599 * 0.664 * 0.568 * 0.594 * 0.636 * 0.508 * 0.567 * 0.633 * 0.623 * 0.672 * 0.756 * 0.420 * 0.432 * 0.508 *														

在 2 尾配对检验下,Exaqt 相对于最强基线 (GRAFT-Net) 的统计显着性用星号 (\*) (< 0.05) 标记。  
表 4:Exaqt 在 TimeQuestions 测试集上与三个复杂 QA 基线的性能比较。

书呆子	回忆#Question实体	
标记我	0.682	2.9
ELQ	0.716	1.7
阿依达	0.541	2.8
TagMe + ELQ 0.758 AIDA + ELQ	3.5	
0.729 TagMe + AIDA 0.701	3.5	
	4.3	

表 5:在测试集上比较各种 NERD 方法。

Exaqt 管道中的步骤	回想一下#候选人	
NERD 实体的所有 KG 事实	0.758	2491
BERT 选择的事实 0.719 为连接注入的最短路径		48
0.720 最大组件上的 GST 0.613 所有组件的 GST 并集 所有组件的完整 GST 由 BERT 添加的时间事实		49
表 6:了解 Exaqt 的面向召回的第 1 阶	0.640	14
段。	0.671	21
	0.724	67

类别	整体显式隐式温度。答。序数				
Exaqt (完整)	0.565	0.568	0.508	0.623	0.420
Exaqt - TCE 0.545 Exaqt - TSE 0.543 Exaqt - TEE	0.556	0.481	0.590	0.406	0.411
0.556 Exaqt - TE 0.553	0.545	0.465	0.598	0.411	0.413
Exaqt - ATR 0.534表 7:检查	0.564	0.475	0.614	0.413	0.398
Exaqt 的面向精度的第 2 阶	0.556	0.495	0.613	0.398	0.411
	0.527	0.465	0.594	0.411	

段。

亚伯拉罕·林肯在担任总统之前做了什么？
特洛伊战争期间谁是特洛伊国王？
哪些影片获得 2009 年奥斯卡最佳影片提名？
内战后哈里特·塔布曼住在哪里？
老板 Bill Neukom 的运动队上次赢得世界大赛是什么时候？

表 8:Exaqt 正确回答的轶事示例。

其他因素提供不同程度的帮助。一个有趣的观察发现,TCE 虽然在大多数情况下发挥着适度的作用类别对于时间性问题最为重要  
答案:即使知道某个问题属于此类帮助模型。  
轶事例子。表 8 显示了测试问题的样本 Exaqt 已成功处理,但没有任何基线。

8 相关工作

IR 中的时间 QA。支持查询中的时间意图文档处理一直是一个长期的研究课题 IR [ 8,14,20,40,49,60 ]。\_\_ 这包括工作内部的具体使用对文本进行 QA 的情况[5,33,46,56 ]。这些努力大部分都需要重要的文档预处理和标记。还有

以特定方式制定问题的责任,以便符合精心设计的解析器。这些方向往往达不到网络上的真实设置,其中包含文档和问题两者都是临时制定的。此外,不幸的是,这种语料库标记在结构化知识图中不起作用。值得注意的时间 QA 方面的工作包括[56] 的工作,它分解将复杂的问题分解为更简单的组件,并将答案片段重新组合为满足原始意图的响应。这样的方法存在解析问题的瓶颈。 Exaqt 不做关于如何提出问题的假设。对 KG 进行时间 QA。具有时间条件的问题有在 KG-QA 文献中没有受到太多关注。为数不多的作品具体解决时间问题的包括[23,38,76 ]。其中, [38]依赖于泛化有限的手工规则,而 Exaqt 则通过远程自动训练

监督并涵盖更广泛的问题领域。 [23] 引入了以事件为中心的 QA 任务,这与我们的任务重叠时间问题的概念,并介绍了基准集合。 [76]提出了一个包含 KG 的键值存储网络有关进入 QA 管道的时间的信息。该方法是针对对于简单的问题,如 WebQuestions 基准测试中所示。 显 KG。最近,将大型 KG 理解为动态的知识体系受到关注,引发了这样的概念 时态知识图或时态知识库[25, 70]。这里,每条边 (对应于一个事实)都与一个相关联时间范围或有效性[43],当前的努力主要集中在关于时间 KG 完成的主题[31,32,42 ]。最近的一个方法探索了对此类时间 KG 的 QA,以及创建相关基准[57]。

9 结论

QA 中的时间问题尚未得到充分探索, KG 中的时间信息也是如此,尽管它们对知识很重要 分析师或记者等工作以及高级信息非专业用户的需求。这项关于 Exaqt 方法的工作提出了填补这一空白的完整流程基于基于 BERT 的分类器和图卷积网络的明智组合。最重要的是,我们设计了用时间信号增强这些组件的新方法。实验结果大复杂时间问题的集合证明了Exaqt 相对于最先进的通用方法的优越性

对知识图进行 QA。

致谢。我们感谢菲利普·克里斯特曼和 Jesujoba 来自 MPI 信息学的 Alabi 在各种方面提供了有用的投入这项工作的阶段。贾震得到 (i) 中国科学院的支持 铁科科学股份有限公司 (2019YJ106) ; (ii) 四川科技计划 (2020YFG0035) 。

参考

[1] Abdalghani Abujabal,Rishiraj Saha Roy,Mohamed Yahya and Gerhard Weikum. 2018.基于知识库的开放领域问答的永无止境的学习.在万维网中。

[2] Abdalghani Abujabal,Rishiraj Saha Roy,Mohamed Yahya and Gerhard Weikum. 2019.ComQA:社区来源的数据集,用于使用释义集群进行复杂事实问答.在 NAACL-HLT 中。

[3] Abdalghani Abujabal,Mohamed Yahya,Mirek Riedewald and Gerhard Weikum. 2017.通过知识图回答问题的自动模板生成.在万维网中。

[4] B Aditya,Gaurav Bhalotia,Soumen Chakrabarti,Arvind Hulgeri,Charuta Nakhe,S Sudarshanxe 等人。 2002.银行:关系数据库中的浏览和关键字搜索.在VLDB中。

[5]大卫·安·史蒂文·肖卡特·马丁·德·科克和艾蒂安·克尔。 2006年。支持临时问答:离线数据收集策略。在CoS-5 中。

[6]詹姆斯·F·艾伦。 1983.维护有关时间间隔的知识。 CACM 26, 11 (1983)。

[7]奥马尔·阿隆索·迈克尔·格茨和里卡多·巴埃扎·耶茨。 2007.论时间信息在信息检索中的价值.在 SIGIR 论坛中。

[8]奥马尔·阿隆索·詹尼克·斯特罗根·里卡多·巴埃扎·耶茨和迈克尔·格茨。 2011年。 时代信息检索:挑战和机遇。 TWAUW 11 (2011)。

[9]加博尔·安杰利·梅尔文·何塞·约翰·普雷姆库玛和克里斯托弗·D·曼宁。 2015.利用语言结构进行开放域信息提取.在

前文续稿。

[10] Sören Auer,Christian Bizer,Georgi Kobilarov,Jens Lehmann,Richard Cyganiak and Zachary Ives。 2007. DBpedia:开放数据网络的核心.在ISWC中。

[11] Junwei Bao, Nan Duan, Zhao Yan, Ming Zhou, and Tiejun Zhao. 2016. Constraint based question answering with knowledge graph. In COLING.

[12]汉娜·巴斯特和埃尔·马·奥斯曼。 2015. Freebase 上更准确的问答.在 CIKM 中。

[13]乔纳森·贝兰特·安德鲁·周·罗伊·弗罗斯蒂格和珀西·梁。 2013.从问答中对 freebase 进行语义解析.在 EMNLP 中。

[14] Klaus Berberich,Srikanta Bedathur,Omar Alonso and Gerhard Weikum. 2010年 满足时间信息需求的语言建模方法.在 ECIR 中。

[15]尼基塔·布塔尼和 HV·贾加迪什。 2019. Hetero在线无模式查询 慷慨的开放知识库.在 CIKM 中。

[16] Nikita Bhutani,Xinyi Cheng,HV Jagadish. 2019.学习通过查询组合回答知识库上的复杂问 题.在 CIKM 中。

[17]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.在 SIGMOD 中。

[18]伯特伦·C·布鲁斯。 1972. 时间参考模型及其在问答程序中的应用.人工智能3 (1972) 。

[19]蔡青青,亚历山大·耶茨。 2013.通过模式进行大规模语义解析 匹配和词典扩展.在ACL中。

[20]里卡多·坎波斯·盖尔·迪亚斯·阿利皮奥·M·豪尔赫和亚当·贾托特。 2014.时代信息检索及相关应用 综述。 CSUR 47, 2 (2014)。

[21] Angel X. Chang and Christopher D Manning。 2012. SUTime:一个图书馆 识别和标准化时间表达.在 LREC 中。

[22] Camille Chaniat,Réduane Dziri,Helena Galhardas,Julien Leblay,Minh-Huong Le Nguyen and Ioana Manolescu。 2018.ConnectionLens:跨异构数据源查找连接.在VLDB中。

[23]塔西奥·苏扎·科斯塔·西蒙·戈特沙克和埃琳娜·德米多娃。 2020.Event-QA:基于知识图的事件 为中心的问答数据集.在 CIKM。

[24] Jacob Devlin,Ming-Wei Chang,Kenton Lee and Kristina Toutanova。 2019.BERT:用于语言 理解的深度双向变压器的预训练.在 NAACL-HLT。

[25] Bhuwan Dhingra,Jeremy R Cole,Julian Martin Eisenschlos,Daniel Gillick,Jacob Eisenstein and William W Cohen。 2021.作为时代知识库的时间感知语言模型.在 arXiv 中。

[26]丹尼斯·迪芬巴赫·佩德罗·恩里克·米利亚蒂·奥马尔·卡瓦斯梅·文森特·卢利·卡迈勒·辛格和皮埃尔· 马雷特。 2019.QAnswer:问答原型,弥合了相当一部分 LOD 云和最终用户之间的差距.在万维网 中。

[27] Bolin Ding, Jeffrey Xu Yu, Shan Wang, Lu Qin, Xiao Zhang, and Xuemin Lin. 2007 年。在数据库中查找前 k 个最小成本连接树.在ICDE。

[28] Mohnish Dubey,Debayan Banerjee,Abdelrahman Abdelkawi and Jens Lehmann. 2019. LC-QuAD 2.0:通过 Wiki 数据和 DBpedia 进行复杂问题回答的大型数据集.在ISWC中。

[29] David Duvenaud,Dougal Maclaurin,Jorge Aguilera-Iparraguirre,Rafael Gómez Bombarelli,Timothy Hirzel,Alán Aspuru-Guzik and Ryan P Adams。 2015.用于学习分子指 纹的图上的卷积网络.在 NIPS 中。

[30]保罗·费拉吉纳和乌戈·斯卡拉。 2010. TAGME:短文本片段的即时注释 (由维基百科实体提供) 。 在 CIKM 中。

[31]阿尔贝托·加西亚·杜兰·塞巴斯蒂安·杜曼契奇和玛蒂亚斯·尼珀特。 2018.学习序列编码器以完成时 间知识图.在 EMNLP 中。

[32] Rishab Goel,Seyed Mehran Kazemi,Marcus Brubaker 和 Pascal Poupart。 2020. 用于时间知识图补全的历时嵌入.在AAAI。

[33] Sanda Harabagiu and Cosmin Adrian Bejan。 2005.基于时间推理的问答.在关于文本问答推理 的 AAAI 研讨会上。

[34]塔赫尔·H·哈维利瓦拉。 2003. 主题敏感的 pagerank:上下文相关的排名 网络搜索算法。 TKDE 15, 4 (2003)。

[35]丹尼尔·埃尔南德斯·艾丹·霍根和马库斯·克罗茨施。 2015 年。具体化 RDF:维基数据什么效果最 好? 。在 SSW@ISWC。

[36]约翰·霍法特·穆罕默德·阿米尔·约瑟夫·伊拉里亚·博尔迪诺·哈根·福斯特瑞·曼·弗雷德·平卡 尔·马克·斯帕尼奥尔·比利亚娜·塔内瓦·斯特凡·塔特和格哈德·韦库姆。 2011.文本中命名实体的稳健消歧.在 EMNLP 中。

[37]贾震,Abdalghani Abujabal,Rishiraj Saha Roy,Jannik Strötgen and Gerhard Weikum. 2018. TempQuestions:临时问答的基准。 在总部。

[38]贾震,Abdalghani Abujabal,Rishiraj Saha Roy,Jannik Strötgen and Gerhard Weikum. 2018.TEQUILA:基于知识库的时态问答。 在 CIKM 中。

[39] Magdalena Kaiser,Rishiraj Saha Roy and Gerhard Weikum。 2021.从知识图对话问答的重构 中强化学习.在西吉尔。

[40]纳蒂亚·坎哈布亚和阿维舍克·阿南德。 2016.时代信息检索.在 一头牛。

[41] Erdal Kuzey,Vinay Setty,Jannik Strötgen and Gerhard Weikum。 2016.随着时间的流逝:具 有时间范围的文本短语的综合标记.在万维网中。

[42]蒂莫西·拉克鲁瓦·纪尧姆·奥博津斯基和尼古拉斯·乌苏尼尔。 2020.用于时间知识库补全的张量分 解.在 ICLR 中。

[43] Julien Leblay and Melisachew Wudage Chekol。 2018.知识图谱中有效性时间的推导.在 TempWeb 中。

[44] Belinda Z. Li,Sewon Min,Srinivasan Iyer,Yashar Mehdad and Wen-tau Yih。 2020.针对问题的高效单端到端实体链接.在 EMNLP 中。

[45] Rong-Hua Li, Lu Qin, Jeffrey Xu Yu, and Rui Mao。 2016. Efficient and progressive group steiner tree search. In SIGMOD.

[46]埃琳娜·洛雷特·赫克托·洛伦斯·帕洛玛·莫塞达·埃斯特拉·萨奎特和曼努埃尔·帕洛马尔。 2011. 文本摘要对语义问答的贡献:在网络上查找答案的新方法.国际智能系统杂志 26, 12 (2011), 1125–1152。

[47]卢晓鲁,Soumajit Pramanik,Rishiraj Saha Roy,Abdalghani Abujabal,Yafang Wang and Gerhard Weikum。 2019.通过将多文档证据与准知识图结合起来回答复杂问题.在西吉尔。

[48]托马斯·米科洛夫·伊利亚·苏茨克韦尔·陈凯·格雷格·S·科拉多和杰夫·迪恩。 2013年。 单词和短语及其组合的分布式表示.在NIPS 中。

[49]科罗拉多州博尔哈·纳瓦罗和埃斯特拉·萨奎特。 2015. 时间合并报告 跨文档事件排序的化和主题建模.在 arXiv 中。

[50] Barlas Oguz,Xilun Chen,Vladimir Karpukhin,Stan Peshterliev,Dmytro Okhonko, Michael Schlichtkrull,Sonal Gupta,Yashar Mehdad and Scott Yih。 2021. UniK QA:开放 领域问答的结构化和非结构化知识的统一表示.在 arXiv 中。

[51]杰弗里·彭宁顿·理查德·索彻和克里斯托弗·曼宁。 2014.GloVe:用于单词表示的全局向量.在 EMNLP 中。

[52] Soumajit Pramanik,Jesujoba Alabi,Rishiraj Saha Roy and Gerhard Weikum。 2021. UNIQORN:基于 RDF 知识图和自然语言文本的统一问答.在 arXiv 中。

[53]詹姆斯·普斯特约夫斯基·贾尼丝·维贝和马克·梅伯里。 2002.多视角和时间问答.在问答中:战略和资 源研讨会计划。

[54]邱云奇,王元卓,金小龙,张坤。 2020.弱监督知识图多关系问答的逐步推理.在 WSDM 中。

[55]里希拉杰·萨哈·罗伊和阿维舍克·阿南德。 2020.通过精选和开放的网络资源进行问答.在西吉尔。

[56] Estela Saquete,J Luis Vicedo,Patricio Martínez-Barco,Rafael Munoz and Hector Llorens。 2009. 增强 QA 系统的复杂时间问题处理能力。 JAIR 35 (2009)。

[57] Apoorv Saxena,Soumen Chakrabarti and Partha Talukdar。 2021.问题 通过时态知识图进行回答.在ACL中。

[58]弗兰克·希尔德和克里斯托弗·哈贝尔。 2003.用于时间问答的时间信息提取.问答的新方向,AAA技 术报告。

[59] Michael Schlichtkrull,Thomas N Kipf,Peter Bloem,Rianne Van Den Berg,Ivan Titov and Max Welling。 2018.使用图卷积网络对关系数据进行建模.在ESWC中。

[60]安德里亚·塞莱。 2002.新闻专线文章中的时间信息:注释方案和语料库研究.博士论文,谢菲尔德大 学。

[61]史宇轩,程功,Evgeny Kharlamov。 2020.通过静态和动态中心标签在知识图谱上进行关键字搜 索.在万维网中。

知识图上的复杂时态问答

CIKM 21,2021 年 11 月 1-5 日,虚拟活动,澳大利亚昆士兰州

[62]詹尼克·斯特罗特根和迈克尔·格茨。2010.HeidelTime:基于规则的高质量时间表达式提取和标准化。在 SemEval 中。

[63] Yu Su, Huan Sun, Brian Sadler, Mudhakar Srivatsa, Izzeddin Gür, Zenghui Yan, and Xifeng Yan. 2016. On generating characteristic-rich question sets for QA evaluation. In EMNLP.

[64] Fabian Suchanek, Gjergji Kasneci 和 Gerhard Weikum。2007.YAGO:语义知识的核心。在万维网中。

[65]海天孙,塔尼亚·贝德拉克斯-韦斯和威廉·科恩。2019.PullNet:通过知识库和文本迭代检索进行开放域问答。在 EMNLP-IJCNLP 中。

[66] Haitian Sun, Bhuwan Dhingra, Manzil Zaheer, Kathryn Mazaitis, Ruslan Salakhutdinov 和 William Cohen。2018。使用知识库和文本的早期融合进行开放域问答。在 EMNLP 中。

[67]孙亚辉,肖晓奎,崔斌,Saman Halgamuge,Theodoros Lappas,罗军。2021.在具有顶点和边权重的图中查找组斯坦纳树。在VLDB中。

[68]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。在 NAACL-HLT 中。

[69] Priyansh Trivedi, Gaurav Maheshwari, Mohnish Dubey 和 Jens Lehmann。2017年。LC-QuAD:通过知识图回答复杂问题的语料库。在ISWC中。

[70] Rakshit Trivedi, 戴汉俊, 王一帆, 宋乐。2017. Know-Evolve:动态知识图的深度时间推理。在ICML中。

[71] Naushad Uzaaman, Hector Llorens 和 James Allen。2012.通过时间问答评估时间信息理解。在国际公务员制度委员会。

[72]斯维特拉娜·瓦库伦科,哈维尔·大卫·费尔南德斯·加西亚,阿克塞尔·波勒雷斯,马尔滕·德里克和迈克尔·科切斯。2019.复杂问题的消息传递和知识图谱的发誓。在 CIKM 中。

[73] Johannes M. van Hulst, Faegheh Hasibi, Koen Dercksen, Krisztian Balog 和 Arjen P. de Vries。2020. REL:一个站在巨人肩膀上的实体。在牛。

[74] Ashish Vaswani, Noam Shazeer, Nick Parmar, Jacob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser 和 Illia Polosukhin。2017 年。您所需要的就是注意力。在 NIPS 中。

[75]丹尼·弗兰德西奇和马斯克·克罗茨施。2014. 维基数据:免费的协作知识库。CACM 57, 10 (2014)。

[76] Wenqing Wu, Zhenfang Zhu, Qiang Lu, Dianyuan Zhang, and Qiangqiang Guo。2020.引入外部知识来回答知识库上具有隐式时间约束的问题。未来互联网 12, 3 (2020), 45.

[77] Mohamed Yahya, Klaus Berberich, Shady Elbassuoni 和 Gerhard Weikum。2013年。通过链接数据网络进行强大的问答。在 CIKM 中。

[78] Ikuya Yamada, Akari Asai, Jin Sakuma, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji 和 Yuji Matsumoto。2020.Wikipedia2Vec:用于学习和可视化维基百科单词和实体嵌入的高效工具包。In EMNLP。

[79]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:使用知识库进行问答。在 ACL-IJCNLP 中。

[80] Xuchao Zhang, Wei Cheng, Bo Zong, Yuncong Chen, Jianwu Xu, Ding Li, and Haifeng Chen。2020. Temporal Context-Aware Representation Learning for Question Routing. In WSDM.