

文章编号: 1003-0077(2022)02-0111-10

基于 Graph Transformer 的知识库问题生成

胡月, 周光有

(华中师范大学 计算机学院, 湖北 武汉 430000)

摘要: 知识库问答依靠知识库推断答案, 需要大量带标注信息的问答对, 但构建大规模且精准的数据集不仅代价昂贵, 还受领域等因素限制。为缓解数据标注问题, 面向知识库的问题生成任务引起了研究者关注, 该任务的特点是利用知识库三元组自动生成问题, 但现有方法仅由一个三元组生成的问题过于简短, 且缺乏多样性。为生成信息量丰富且多样化的问题, 该文采用 Graph Transformer 和 BERT 两个编码层来加强三元组多粒度语义表征以获取背景信息, 在 SimpleQuestions 数据集上的实验结果证明了该方法的有效性。

关键词: 问题生成; 知识库; 语义表征; 知识库问答

中图分类号: TP391

文献标识码: A

Question Generation from Knowledge Base with Graph Transformer

HU Yue, ZHOU Guangyou

(School of Computer, Central China Normal University, Wuhan, Hubei 430000, China)

Abstract: Knowledge base question answering requires a large number of question answering pairs. To alleviate the problem of data annotation, the question generation from knowledge base has attracted the attention of researchers. This task is to use the triples of knowledge base to automatically generate the questions. To generate questions with rich and diverse information, this paper uses two encoding layers, Graph Transformer and BERT, to enhance the multi-granular semantic representation of triples to obtain background information. Experimental results on the SimpleQuestions dataset prove the effectiveness of the method.

Keywords: question generation; knowledge base; semantic representation; knowledge base question answering

0 引言

问答 (Question Answering, QA) 系统利用信息检索, 在海量的非结构化或者结构化的数据中推断自然语言问句的答案。问答任务是人工智能的核心研究之一, 就目前所研究的 QA 而言, 大部分的训练数据都是以 QA 对作为标注数据, 例如, 基于知识库的 SimpleQuestions^[1] 数据集, 基于文本的 Wiki QA^[2] 数据集, 以及阅读理解所用的 SQuAD^[3] 数据集等。这些数据集受规模大小等条件限制, 且标注花费昂贵, 随着问答系统、阅读理解等领域的发展, 对标注数据的需求越发强烈, 因此, 本文探索的问题生成 (Question Generation, QG) 任务可以为 QA 系统的研究提供一个扩充数据集途径。

当下, QG 任务开始逐渐成为一支有力的研究方向。给定一篇包含多个事实/知识点的文本和标准答案, 传统的 QG 任务根据答案与这些事实逆向生成内容丰富、角度多元化的问句。随着深度学习的推进, QG 任务也开始具有了多样性。给定一些数据, 例如, 结构化的知识图谱或 SQL, 半结构化的表格, 甚至无结构化的文本等, QG 系统都能根据输入的数据自动生成自然语言的问题。问题生成虽然不是新兴任务且其研究领域较为小众, 但随着图像处理、自然语言处理等各领域研究的相互影响, 使得研究者们开始将注意力放在了 QG 任务上。基于知识库的 QA 系统很难理解自然语言问题, 依靠目前常用的知识图谱和数据库也不一定可以回答出该问题。由于 QG 系统可以从已有的知识库生成 QA 系统需要的标准“问题-答案”对数据, 故当出现

收稿日期: 2021-02-21 定稿日期: 2021-03-26

基金项目: 国家自然科学基金 (61972173)

一个新问题时, QG 系统生成的问句与已有的“问题-答案”对进行相似计算, 向用户推荐知识库内已存在的相接近的问题, 间接解决问答的问题。此外, 由于能够与 QA 任务共用数据集, 如给定知识库时, 生成的问句可以为基于知识库的问答系统提供更大的训练数据集, 与基于知识库 QA 任务的可成为对偶任务。传统的 QG 系统使用人工设定的模板或规则来生成问答对解决特定领域的 QA 任务。本文着重于结构化知识库上的 QG 任务。传统的问题生成方法多采取模板的方式。例如, Duma 和 Klein^[4] 使用关系定义的模板来生成简短的描述, 并相应地替换主语和宾语实体的占位符标记。另外 Seyler 等人^[5] 从知识库三元组中生成问句, 实体与谓词的表达由它们在知识库和给定词典中所存在的标签决定。Seyler 等人^[6] 参考基于模板的方法来描述结构化查询并生成自然语言问题。然而, 这种基于规则的方法无法识别单词的语义内容, 且可扩展性较弱。

如今随着深度学习的热潮, 尤其是随着 Sequence-to-Sequence^[7] (Seq2Seq) 框架和编码-解码结构^[8] 为自然语言生成^[9] 任务带来了新的研究方向, 并在机器翻译^[10]、智能对话^[11] 等任务获得优异的成绩, 一系列以基于神经网络的 QG 系统被提出。Serban 等人^[12] 采用基于注意力的编码-解码框架在 SimpleQuestions 数据集上训练模型, 并生成了 30MB 的标签数据。Khapra 等人^[13] 将给定知识库中的所有实体转换为一组关键字, 然后以 Seq2Seq 结构进行建模。ElSahar 等人^[14] 采用 Zero-Shot 使模型能够泛化未遇见过的谓词和实体的问题。Wang 等人^[15] 在 Seq2Seq 基础上加入复制机制^[16], 使模型能解决 OOV 问题。

上述生成模型只考虑到对未出现过的三元组或低频出现的生僻词的解决方法, 但对于生成信息丰富和多样化的问题的考虑有所欠缺。受 Cai^[17] 和 Koncel-Kedziorski 等人^[18] 工作的启发, 本文着重于加强对三元组的多粒度语义特征表示, 采用双编码层: 基于 Graph Transformer 的图编码层和基于 BERT^[19] 加强的词级编码层。本文预先将知识库中的实体、关系构成知识图谱, 赋予实体全局化的语义向量, 然后针对该三元组, 结合 Transformer 结构^[20] 的并行性对输入节点进行特征细化。同时, 为了充分利用词语粒度的语义向量, 三元组的词语序列先通过 BERT 预训练模型获取向量表征, 再使用双向门控循环单元 (Gated Recurrent Unit, GRU)^[8] 网络计算上下文向量。最后, 本文将两个编码层联合获得更完善的三元组特征表示, 再输入解码层生成问句。本文在英文数据集 SimpleQues-

tions 上进行实验, 实验的评测结果证明了该方法的有效性。综上, 本文的贡献如下:

(1) 基于知识库的 QG 模型, 率先提出使用基于 Graph Transformer 的方法为实体获取丰富的背景信息;

(2) 为完善三元组的特征表示采用图和词级的表示, 结合知识图谱和 BERT 分别进行初始化, 再用神经网络对其泛化;

(3) 自动评测和人工评估的结果表明, 本文模型生成的问题信息量更丰富, 表达形式更多样化。

1 模型设计

为了解决知识库中单个三元组存在背景信息量少、语义表达不够完善等问题, 受 Koncel-Kedziorski^[18] 和 Cai^[17] 等人的工作启发, 本文采用具有图神经网络特征的 Transformer 结构作为编码层, 结合知识图谱对实体和关系进行表示, 以图的形式作为输入, 并且还使用 BERT 预训练模型获得词语的语义表示。与 Seq2Seq 模型结合, 从而针对任务目标规范适合本数据集的语义向量表示, 以获得更准确的潜在语义, 使得模型生成的问题更加丰富流畅。

1.1 模型结构

本节将介绍基于 Graph Transformer (GT) 的知识库问题生成模型 (简称 GT-KBQG)^[20-21]。GT-KBQG 模型的结构如图 1 所示, 该模型由三个

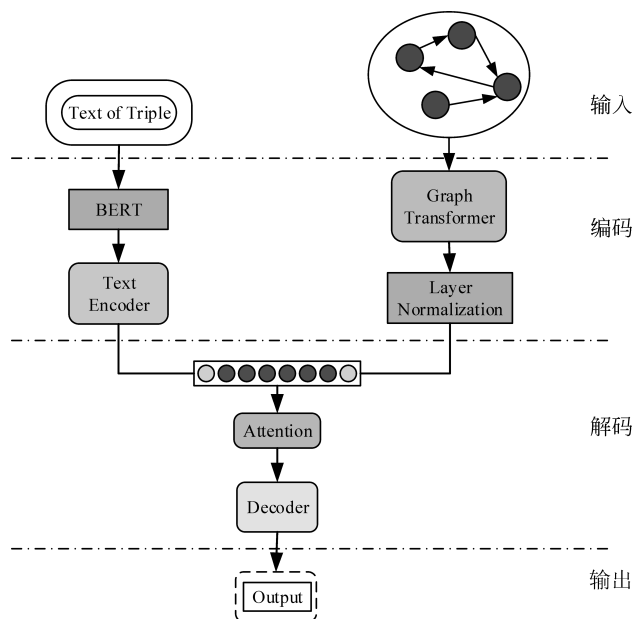


图 1 模型结构图

双编码层在解码端连接

模块组成,分别为 GT 编码器、基于 BERT 预处理的文本编码器,以及基于 Attention 机制的 GRU 解码器。具体来看,首先三元组进行构图预处理,之后使用 GT 编码器对三元组在图中的向量表示 $e_{\text{triple}} = (e_s, p_r, e_o)$ 进行编码,同时以自然语言文本形式表示的三元组 $w = \{w_1, \dots, w_i, \dots, w_m\}$ 通过 BERT 预处理转换为向量形式 $x_w = \{x_{w_1}, \dots, x_{w_i}, \dots, x_{w_m}\}$,通过双向 GRU 编码器编码。然后将以上两个编码层结果结合作为解码器的隐藏层初始化状态,解码器结合 Attention 机制^[10]获得上下文向量,以此加强问题的生成效果。

1.2 基于 Graph Transformer 的图编码层

为了让实体之间的联系更加紧密和结构化,有学者将其构成图的形式(实体为节点,关系为边),即知识图谱。本文利用已构成的知识图谱赋予数据集中的三元组丰富的背景知识,结合 Transformer 结构,在丰富的背景知识中抓取对本数据集针对任务有利的语义信息,使实体之间获得更紧密的关系,使单个实体的语义向量也更加丰富。

1.2.1 知识图谱预处理

英文知识库 FreeBase^[22] 的实体原始表示为独立的 id 形式,且 FreeBase 主要由社区成员收集整理,数量庞大,常被视为可靠的知识图谱。SimpleQuestions 数据集基于 FreeBase 得到,故其原始实体也是 id 形式,为了使模型能够获取更多的实体背景信息,本文将以大规模知识库预先训练获得的知识图谱作为 SimpleQuestions 数据集中实体的信息网背景图。与词嵌入的思想一致,本节主要是获得知识图谱的嵌入(Knowledge Graph Embedding, KG-Emb)^[23]。

KG-Emb 是为了将知识图谱中的高维的谓词、实体表示为低维的形式(P 和 E)。为实现这一目标,Bordes 等人^[23]提出基于翻译模型的 TransE,通过训练 P 和 E 两个矩阵,使所有事实(s, r, o)的总距离 $\sum \|e_s + p_r - e_o\|_2^2$ 最小。在 TransE 的推动下,探索了一系列基于翻译的模型。例如,由 Wang 等人^[24]提出的 TransH 处理一对多关系或者多对一关系,与 TransE 直接度量 e_s 和 e_o 之间的距离不同,TransH 将其投影到谓词特定的超平面中,在一定程度上解决了 TransE 不善于处理复杂关系的情况,使同一个实体在不同关系下的表示不同。TransH 预设实体与关系处于相同的语义空间,Lin 等人^[25]提出的 TransR 则将关系处于不同的语义空

间的假设,该方法为每个谓词 r 定义一个转换矩阵 M_r ,以最小化 $\sum \|e_s M_r + p_r - e_o M_r\|_2^2$ 为目标。基于 TransE 进行改进的类似算法还有很多,比如 PTransE^[26] 等多种图嵌入表示方法。

本文选用 TransE 方法对事实三元组进行图的预处理,将实体作为整体转为向量的表示。选择 TransE 主要有以下几个考虑:首先,与其改善结构如 TransH、TransR 等相比,TransE 的参数较少;其次,TransE 的主要层次关系的表示非常有效;最后,所有关系模型在多关系的实验分析中,TransE 的效果良好。本文直接采用了 Huang^[27] 等人在 FB2M 数据集上所提供的基于 TransE 的 KG-Emb。

1.2.2 Graph Transformer

本文使用预处理后的知识图谱作为 KG-Emb,本编码层的输入是具有全局语义信息的实体和关系,并基于 Transformer 结构,在全局图的背景语义下进一步捕捉主语实体、关系、宾语实体之间的关联,使作为输入的三元组具有更加符合本任务的语义粒度表示。Vaswani 等人^[21]提出的 Transformer 结构,通过全局上下文建模的多头自注意力机制来实现高效且并行的计算,具有并行性的优点,克服了 RNN 的在长序列上的顺序计算结构的缺点。

作为图编码器,GT 参考 Transformer 编码层结构,由几个相同的网络块组成,根据其并行计算特性,输入节点之间可直接进行信息传递,图 2 表示的为单个网络块,左边部分的 e_s 、 e_o 和 p_r 分别表示从 KG-Emb 中获取的主语实体、宾语实体以及谓词的语义向量,通过 Graph Transformer 结构计算抓取语义信息,使实体向量更加丰富和符合本文任务。

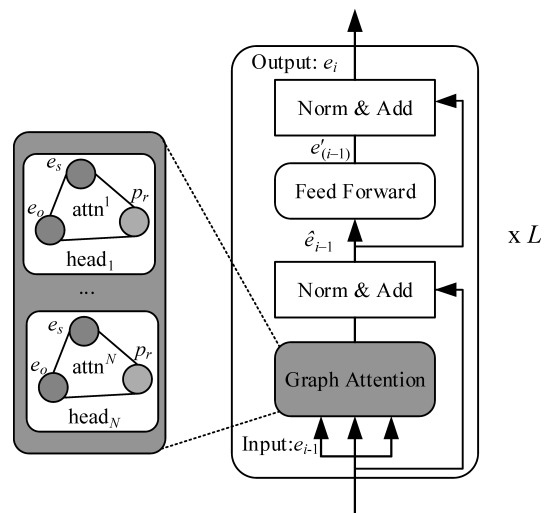


图 2 Graph Transformer 结构

本文使用 N 头注意力机制在图 2 所示的网络块中堆叠 L 次,在输入残差网络之前进行拼接。Graph Transformer 的具体计算如式(1)所示。

$$\hat{e}_i = e_i + \left\|_{j=1}^N \text{attn}^j(q_i, k_i) v_i \right. \quad (1)$$

其中, $\|$ 表示 N 个 attn 的连接操作($j \in N$), attn^j 为点积计算。如式(2)所示,为了降低点积造成的阻碍梯度流趋势,参照 Vaswani 等人^[21]的方法,将结果缩小 $\sqrt{d_k}$ 之后再利用 softmax 函数进行归一化。

$$\text{attn}^j(q_i, k_i) = \text{softmax}\left(\frac{q_i k_i^T}{\sqrt{d_k}}\right) \quad (2)$$

式(1)和式(2)中出现的 q_i, k_i, v_i 是第 i ($i \leq L$) 个堆叠块对输入进行线性变换后的 d_k 维的向量表示,如式(3)所示。

$$\begin{cases} q_i = e_i W_i^q \\ k_i = e_i W_i^k \\ v_i = e_i W_i^v \end{cases} \quad (3)$$

其中, e_i 由第 $i-1$ 层标准化计算得来,即为第 i 层的输入。通过块状网络扩展多头注意力层,如图 2 所示,每个块状网络都具有衔接层转换,如式(4)、式(5)所示。

$$e'_i = \text{FF}(\text{LN}(\hat{e}_i)) \quad (4)$$

$$e_{i+1} = \text{LN}(e'_i + \text{LN}(\hat{e}_i)) \quad (5)$$

式(5)中,第 i 层的最终输出作为第 $i+1$ ($\leq L$) 层的输入。式(4)中, $\text{FF}(\cdot)$ 为两层前馈网络,其中, LN 为标准化层,两层之间具有非线性变换,本文选取 ReLU 非线性变换以及残差网络的前馈进行计算。通过知识库构图之后,每个实体和关系拥有对应的向量化表示,实体与关系的拼接作为 $i=1$ 时的初始输入,如式(6)所示。

$$e_1 = \text{Concat}(e_s; p_r; e_o) \quad (6)$$

图 1 中的 Layer Normalization 层,对最终的输出结果 e_N 实行层标准化如式(7)所示,与词级编码层进行融合。

$$e_N = \text{LN}_{\text{output}}(e_N) \quad (7)$$

多个块状网络的堆叠使得信息可以从图进行传播。从 GT 编码层输出的最终结果与下一节的文字编码结果进行组合作为解码器端的输入。

1.3 基于 BERT 增强词级表示的编码层

在 2018 年, Jacob Devlin 等人^[19]提出 BERT 预训练模型,在自然语言处理领域获得极大的反响,在预训练和精调下,刷新了十多项 NLP 任务的记录。

BERT 是基于 Transformer 结构的 Encoder,其并行结构擅长捕捉长距离依赖,避免了 RNN 的顺序结构带来的缺点。BERT 为双向 Transformer 结构,未标记文本的深层双向表示由所有层的双向上下文共同进行条件化来预训练。与已提出的预训练模型相比,经过巨量语料训练的 BERT 具有捕捉到真正意义上的双向上下文信息的能力。由于 BERT 的出色表现,在 Python 库中已经收纳整理了可以调用的 BERT 函数库,使用者安装好 BERT 环境^①,再自行将所需数据导入就可以直接使用。SimpleQuestions 数据集中实体原始为独立的 id 形式,而在目标问题中用自然语言的单词形式表示,且 id 形式尽管可以展现出实体的唯一性,但实体的描述词语中会出现一些语义相通的词汇,有助于问题的生成,所以本节将三元组的实体以自然语言形式作为模型输入的一部分。本文主要考虑通过表现良好的语义表示来生成问句,以此来验证这些表示是否在本任务中有效。考虑到 BERT 在编码方面表现优异,本文用其捕捉词语级别的表示。

由上文可以了解到,Transformer 擅长捕捉长距离信息,适用于长文本任务。而三元组的词级输入为短字符串,采用顺序计算结构的 GRU 作为编码器更有优势,通过 GRN 双向编码,获得三元组词语粒度的嵌入向量表示。首先,通过 BERT 预处理泛化每个词的嵌入向量的语义信息,再采用双向 GRU 的最后一个隐藏状态作为每个词生成 d 维嵌入,如图 3 所示。

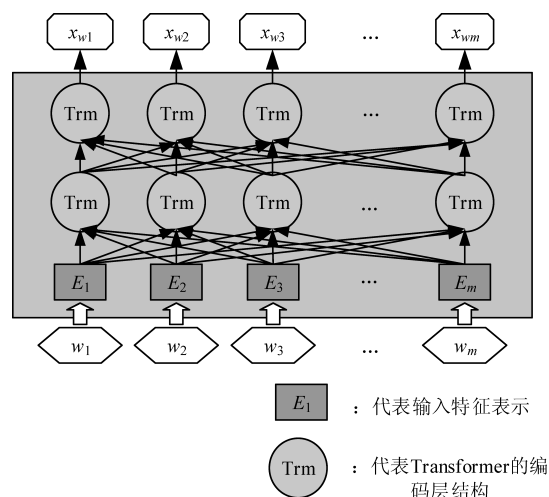


图 3 BERT 预处理结构图

① <https://github.com/hanxiao/bert-as-service>

图 3 中,每个输入三元组表示为单词型序列 $w = \{w_1, \dots, w_i, \dots, w_m\}$,调用 BERT 函数进行预处理编码,得到维度一致的向量矩阵 $x_w = \{x_{w1}, \dots, x_{wi}, \dots, x_{wm}\}$ 。然后,将经过 BERT 预处理后的语义向量输入文本编码器中,即图 3 的词级编码器 (Text Encoder),其由双向 GRU 对 x_w 进行正向和反向读取来抓取输入序列 x_w 的有效语义信息,详细结构如图 4 所示,式(8)和式(9)可表示该过程。

$$\vec{h}_i = \text{GRU}(\vec{h}_{i-1}, x_{wi}) \quad (8)$$

$$\overleftarrow{h}_i = \text{GRU}(\overleftarrow{h}_{i+1}, x_{wi}) \quad (9)$$

其中, \vec{h}_i 和 \overleftarrow{h}_i 分别为正向与反向的隐藏状态表示,再将每一步的正反隐藏状态拼接,得到整合的隐藏层状态 $h = \{h_1, \dots, h_i, \dots, h_m\}$ 作为最后的编码层输出,其中, $h_i = [\vec{h}_i; \overleftarrow{h}_i]$ 。

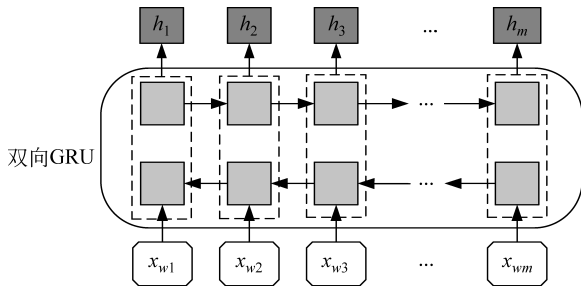


图 4 Text Encoder 结构图

本文将 e_N 和 h 拼接作为多编码层的最终输出 $\tilde{h} = [e_N; h; e_N]$ 。选择图编码层的最后一层的输出 e_N 和 Text Encoder 的最后一层隐藏状态 h_m 经过式(10)成为解码层的隐藏状态的初始化值。

$$s_0 = \text{ReLU}([e_N; h_m]W_s) \quad (10)$$

1.4 基于 Attention 机制的 GRU 解码层

本文的解码器用来预测生成语义上流利且合适的简单问题,由于 Transformer 的并行结构性能效果非常好,我们观察了在 Transformer 结构上的任

务,发现优良的并行结构使得模型在计算时存在注意力过长的问题,导致了生成的序列中出现了与输入不相关的词汇。

由于本文任务生成简单问题而非长序列,顺序结构的 RNN 类网络每步的预测计算都依赖上一步的输出,其过程更贴合本文的任务目标。因此,本文使用基于注意力机制的 GRU 作为解码器。生成第 t 个单词的概率分布如式(11)所示。

$$p(y_t | y_{\leq t}, x_w) = \text{softmax}(W_o \bar{s}_t) \quad (11)$$

其中, \bar{s}_t 是经过注意力机制之后的 t 时刻隐藏状态,其由上下文向量 c_t 和 GRU 计算得到的隐藏状态 s_t 进行计算得到,如式(12)所示。

$$\bar{s}_t = \tanh(W_a [c_t; s_t]) \quad (12)$$

当前 t 时刻的隐藏状态 s_t 和通过注意力机制计算上下文向量 c_t 的详细计算由式(13)和式(14)给出。

$$s_t = \text{GRU}(y_{t-1}, c_{t-1}, s_{t-1}) \quad (13)$$

$$c_t = \sum_{r=1}^{m+2} \alpha_r \tilde{h}_r \quad (14)$$

其中, α_t 为第 t 时间步的对齐向量,详细计算如式(15)所示。

$$\alpha_t = \frac{\exp(s_t^T \tilde{h})}{\sum_{r=1}^{m+2} \exp(s_t^T \tilde{h}_r)} \quad (15)$$

为了保证生成问句时句子主语的完整性,本文采用模板方法,将主语用占位符表示,在完成预测之后,再将问题对应的三元组主语补充上去。

2 实验结果与分析

2.1 实验数据

本文使用了具有 id 形式的英文数据集 SimpleQuestions,表 1 为文本与 id 形式的问答对示例。本文的输入分为两部分。

表 1 SimpleQuestions 数据集 id 和文本形式样例

主语实体	谓词	宾语实体	目标问题
id: m/0ms5mg text: most of us are sad	Music/recording/artist	id: m/0mjn2f text: Eagles	Which artist recorded most of us are sad?
id: m/086k8 text: warner bros entertainment	Film/production_company/films	id: m/0278x5r text: Saving Shiloh	What movie is produced by warner bros.?
id: m/02dtg text: Detroit	Location/location/people_born_here	id: m/01s8mcb text: J. Moss	Who is a musician born in Detroit?

其一为对三元组进行构图预处理之后的实体-关系的向量表征。在实验过程中,对输入三元组进行构图操作,FreeBase 被视为可靠的知识库,SimpleQuestions 中所有事实基于 FreeBase。由于 FreeBase 规模巨大,故在一些研究工作中,对 FreeBase 实行子集提取,如 FB2M、FB5M 等。SimpleQuestions 中的三元组可在 FreeBase 的子集 FB2M 里查询到,故本文选取 Huang 等人^[27]所提供的基于 FB2M 训练 TransE 得到图嵌入作为本文 KG-Emb。表 2 列出了 FB2M 与 SimpleQuestions 两者的具体信息。

表 2 数据集统计信息

#	FB2M	SimpleQuestions
Training	14 174 246	75 910
Validation	N.A	10 845
Test	N.A	21 687
Predicates	6 701	1 837
Entities	1 963 130	131 681
Vocabulary	733 278	61 336

其二为自然语言形式下的文本三元组通过 BERT 预处理之后的词向量表征。BERT 预处理模型经过海量数据的预训练,可以赋予词语更完善的语义向量。并且,在编码端无需人为构造单词表,所有单词都可由 BERT 获得语义信息丰富的向量表征,这可以减轻诸多实体里出现的大量低频词汇所带来的单词量的压力,以加快模型的计算速度。

2.2 评价标准

本文使用了一组在自然语言处理领域中较为完善的自动评估指标来评测模型效果: BLEU^[28], METEOR^[29], ROUGE-L^[30]。考虑到语言表达的复杂性,本文还采用了人工评测的方法,邀请三位自然语言处理领域的研究生对生成的问题进行评估。

BLEU: BLEU 在翻译任务中最早被使用,是将候选文本与一个或多个参考文本进行比较的评估标准。它是一种语法度量,用于计算生成的文本与参考文本之间的 n 元重叠。由于 BLEU 只考虑单词的连续匹配准确率,故其无法考虑到句子的语法以及语义方面的效果,如 BLEU-1 考虑的是单个单词之间的匹配度, BLEU-2 则考虑连续两个单词的匹配度等。

METEOR: METEOR 是基于准确率与单词召回率上的加权调和平均数,其目的是改进 BLEU 标

准固有的一些缺点。它还考虑到了一些其他指标没有发现的功能,例如,同义词匹配等,使其与人工评价的效果距离更近一步。

ROUGE-L: ROUGE 是使用基于召回率的相似度量进行计算的指标,其基本思想是使用模型生成的文本和参考文本的 n 元组共现概率作为评价的基础,但无法评价句子是否流畅。在本文的评估中,选择 ROUGE-L 作为评估标准,其基本思想是匹配两个文本单元之间最长的公共序列。

人工评估 (Human Evaluation): 尽管自动评测在某种程度上可以估量生成的问句与参考问句有多相近,但是当语义相近而语法表达结构不同时,许多隐藏着的限制就显示出来。由一个目标问句和两个模型预测的问句组成的一个例子可以更清楚地说明这个问题,如表 3 所示。尽管表中的三个句子在语义表达方面类似,但是利用自动评测标准来度量时分数差距却有将近 20 个百分点。此外,自动评测也不能确定是否有三元组中相关的关系词出现在预测的问句中。因此,人工来评测给出句子的流畅度与相关度是必不可少的。

表 3 BLEU 值得分举例

目标语句	Who directed walter hill filmmaker ?	BLEU
生成句 A	Who was the director of walter hill ?	58.06
生成句 B	Who directed the film walter hill ?	77.58

注:表中 BLEU 分数值为 BLEU-1 至 BLEU-4 的得分均值。

2.3 基准模型

本文做了多个对比实验以证明提出方法的有效性,具体的对比模型如下详释,Zero-Shot 与 Enc-Dec 模型的计算以及预处理保持与原始文章一致,其余模型保持与本文一致的预处理,但不加入图层次的编码层。

Transformer^[21]: Transformer 是当前在机器翻译任务上非常火的模型。本文使用 Transformer 作为一个对比实验,观察它在此文任务中的表现。

Zero-Shot^[14]: Zero-Shot 利用特殊标识符替代关系词以解决在测试集可能出现的而在训练集里未出现的关系词的问句。在该模型中,SimpleQuestions 数据集三元组的实体和关系以 TransE 嵌入为初始化。

Enc-Dec^[12]: Enc-Dec 是根据 Serban 等人的工作来实现的。本文只使用了单占位符, Serban 等人还使用了多类占位符 (MP) 作为输入序列,但由于

在文章中并没有将这些占位符的类别罗列出来,并且 MP 所带来的贡献并不大,故在本文的实验方法中,不能报告以 MP 作为输入的实验结果。

CopyNet: Wang 等人^[15]在基于知识库的 QG 任务上融入了复制机制,本文复现了融合复制机制的 CopyNet 模型。

2.4 实验设置

本文使用多编码层,在单词粒度上的编码器使用双向 GRU 结构,每个 GRU 隐藏节点数为 500,由 BERT 初始化词向量,维度为 768;基于图的编码器使用 5 个多头注意力机制和 4 个循环块的 Transformer 结构,使用 TransE 对三元组进行构图初始化,维度为 250。解码器使用结合 Attention 的单层 GRU 结构,每个 GRU 隐藏节点数为 1000,用 GloVe^[31]预训练好词嵌入对词向量初始化。本文基于 Tensorflow 平台^[32],采用 Adam 优化算法^[33]。学习率初始化为 $1e-3$,每 500 步进行一次衰减。模型使用 mini-batch 方式训练,batch 大小设置为 64。

2.5 结果分析

2.5.1 自动评测

如表 4 所示,GT-KBQG 在 BLEU-1、BLEU-2、METEOR 和 ROUGE-L 指标上性能提升效果明显。Transformer 模型虽然具有并行性的计算优势,但仅靠输入的单个三元组获取的语义背景信息

始终不够完善,依赖并行结构的特征,其得到的效果虽然在 BLEU-1 上优于 Zero-Shot 和 Enc-Dec 模型,但在 BLEU-2、BLEU-3 和 BLEU-4 指标上均逊色于 Zero-Shot 模型。与 Enc-Dec 模型相比,GT-KBQG 在 METEOR(+2.98)和 ROUGE-L(+9.96)指标上性能提升非常明显,从一方面说明 GT-KBQG 生成的问题与目标问题的语义和词语共现效果更好,但在 3-gram 和 4-gram 重叠(BLEU-3、BLEU-4)匹配上,效果却不明显,说明了词语共现性提升,但可能表达形式更丰富,故在多元匹配上与目标问句相符的较少。与其他基准模型对比,CopyNet 模型效果提升也十分明显,但由于其着重于对三元组中出现的低频词进行处理,而未考虑信息表达的多样化,故在 1-gram 和 2-gram 上的词语共现不如 GT-KBQG,且在 BLEU-3 和 BLEU-4 上的效果也逊色于 Zero-shot 模型。虽然在 BLEU-1 的匹配上,GT-KBQG 效果最优,但是在多元匹配上与 Zero-Shot 依然存在差距。根据后面的人工评价,我们推测是因为与参考问句的一致性导致的。多元匹配对词语组成问句序列具有严格要求,但由于 GT-KBQG 模型的 BLEU-1 效果最好,又说明本模型具有与参考问句相匹配的更多个词语,推测是因为在语句表达形式上与参考问句一致性没有 Zero-Shot 高,但在总体的语义表达上,METEOR(+1.81)和 ROUGE-L(+8.92)又优于 Zero-Shot,说明虽与参考问句表达形式一致性不高,但是表达目的应该是一致的。

表 4 自动评估结果

#	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L
Transformer	68.56	43.31	31.32	23.29	36.31	68.80
Zero-Shot	62.44	50.62	40.82	31.10	36.24	61.32
Enc-Dec	60.92	46.05	36.32	27.36	35.07	60.28
CopyNet	69.74	46.28	34.92	27.12	37.48	69.80
GT-KBQG	72.54	47.37	35.81	27.86	38.05	70.24

2.5.2 人工评价

由于语言表述的复杂性,自动评测只能表现出针对单一目标问题的参考文本的效果,无法完全说明生成的问题是否合适。因此本文还以人工评价的方式对模型进行评估。表 5 中 Sim.(similar)表示生成的问句与参考问句近乎一致的句子占被选评估集的比例,基线模型与本模型所占比例都达到一半左右,具有较高的一致性。本文除了要考虑生成问句的一致性外,还关注生成句子是否具有多样性,同

时对背景信息描述是否丰富,表 5 中的 Var.(variety)表示生成的问句比参考问句描述更丰富的比例。

本文在测试集中随机抽取三份数据用于人工评价,每份数据集有 300 个三元组-问题对,同时每份数据集以 3 个人参与,与标准问题以一对示例的形式进行标注。由表 5 所示,GT-KBQG 模型与目标问题的一致性不是最高,即在 Sim. 结果上,比 Transformer(-0.99%)、Zero-Shot(-2.47%)、Enc-Dec(-1.42%)和 CopyNet(-2.4%)的结果都

表 5 人工评估结果

(单位: %)

#	Transformer	Zero-Shot	Enc-Dec	CopyNet	GT-KBQG
Sim.	49.86	51.34	50.29	51.27	48.87
Var.	42.90	40.72	38.72	43.76	57.43

低,可以推测出目标问题比较简单,但也会出现不少具有实体描述信息的问句。但是在 Var.上的评价结果 GT-KBQG 最好,比基准模型中表现最好的 CopyNet 高了 13.67 个百分点,这证明了本文对三元组特征表示的加强赋予模型更加丰富的背景信息,从而使生成的问句中的实体描述的信息更多。

2.5.3 样例分析

在评测的结果上,只能基于评估数据对生成的问题进行推测分析,无法直观感受生成问句的质量效果,表 6 选取目标问题与 Enc-Dec 模型 GT-KBQG 模型进行样例对比。

根据表 6 中所展示的样例可以看出,比起 Enc-Dec 模型,GT-KBQG 能生成更加具有描述性的问题,如在示例 1 和示例 3 中,Enc-Dec 所生成的问题正确但非常简洁,实体描述信息非常少,该类问题可以有非常多的回答,若与 QA 系统作为对偶学习任务,可能会降低系统的准确度。相比以上两个模型,GT-KBQG 对于提问对象“Robert Drummond”

的描述更进一步,该问题提供了“Robert Drummond”的之前的职业信息“football player”,比 Human 对实体所述的信息更细,但比起人工生成的问题,GT-KBQG 的描述仅针对了对象实体,Human 行对主题实体也进行了描述,可以让读者了解“Syracuse”处于 New York。但总体上来看,与 Enc-Dec 模型相比,GT-KBQG 生成问句时着重于对实体背景信息的描述,产生的问题更加精准化和多样化。在自动评测结果中,尽管对 Enc-Dec 模型进行了改善,但在多元重叠效果上并不太明显,对生成的问题进行人工评价,发现其原因除了模型自身依旧存在缺陷之外,与目标问题也有关联。例如,样例 2,目标问题较为简单,Enc-Dec 模型本身存在生成句简短的特征,故与目标问题完全匹配,而进行改善的模型尽管生成了更加丰富的实体背景信息,但在自动评测指标上的结果却不如 Enc-Dec 效果好,而 GT-KBQG 预测出的背景信息越多、越完善,其 BLEU 等自动指标结果值就会越低。

表 6 样例对比

#	三元组	模型	生成的问题
1	-Syracuse- -people born here- -Robert Drummond—	Human	Which football player was born in Syracuse New York?
		Enc-Dec	Who was born in Syracuse?
		GT-KBQG	Which former professional football player was born in Syracuse ?
2	-lady penelope creighton-ward- -character created by- -gerry anderson—	Human	Who created lady penelope creighton-ward?
		Enc-Dec	Who created lady penelope creighton-ward?
		GT-KBQG	Who created the fictional character lady penelope creighton-ward?
3	-marcus allen- -notable types- -american football player—	Human	which american sport does marcus allen play in?
		Enc-Dec	what is marcus allen?
		GT-KBQG	which sport is marcus allen known for?

3 总结与展望

本文提出了基于 Graph Transformer 的问题生成模型,缓解了模型生成单一化的问题。与直接附加上下文信息作为输入的模型不同,本文仅对三元

组本身的表示进行多样化加强,进而获得丰富的语义背景信息。在 SimpleQuestions 数据集上进行实验,与基准模型相比,人工评测与自动评测的结果证明了对三元组语义加强操作在一定程度上丰富了问题生成的多样性。本文尽管在实验结果上展现出较好的效果,但依旧有许多问题待解决,比如与人工生

成的问题依旧存在差距,且遇到未涉及的三元组或实体关系时,需要一个更完善的图。在后续工作上,可以考虑构造出更大、更完善的知识图谱。由于时间、环境、设备、页面篇幅等因素的限制,本文提出的 GT-KBQG 模型只在 SimpleQuestions 数据集上进行了实验验证,未来的工作将会在多种数据集上进行研究,并将进一步进行消融实验分析。

参考文献

- [1] Bordes A, Usunier N, Chopra S, et al. Large-scale simple question answering with memory networks[J]. arXiv preprint arXiv: 1506.02075, 2015.
- [2] Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78-85.
- [3] Rajpurkar P, Zhang J, Lopyrev K, et al. SQuAD: 100,000+ questions for machine comprehension of text[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2016: 2383-2392.
- [4] Daniel Duma, Ewan Klein. Generating natural language from linked data: Unsupervised template extraction[C]//Proceedings of the IWCS, 2013: 83-94.
- [5] Seyler D, Yahya M, Berberich K. Generating quiz questions from knowledge graphs[C]//Proceedings of the 24th International Conference on World Wide Web, 2015: 113-114.
- [6] Seyler D, Yahya M, Berberich K. Knowledge questions from knowledge graphs[C]//Proceedings of the ICTIR, 2017: 11-18.
- [7] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Proceedings of the Advances in Neural Information Processing Systems, 2014: 3104-3112.
- [8] Cho K, van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2014: 1724-1734.
- [9] 张建华, 陈家骏. 自然语言生成综述[J]. 计算机应用研究, 2006, 23(8): 1-3.
- [10] Luong T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015: 1412-1421.
- [11] 贾熹滨, 李让, 胡长建, 等. 智能对话系统研究综述[J]. 北京工业大学学报, 2017, 043(009): 1344-1356.
- [12] Serban I V, García-Durán A, Gulcehre C, et al. Generating factoid questions with recurrent neural networks: The 30M factoid question-answer corpus [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 588-598.
- [13] Mitesh M Khapra, Dinesh Raghu, et al. Generating natural language question-answer pairs from a knowledge graph using a RNN based question generation model[C]//Proceedings of the EACL, 2017: 376-385.
- [14] Elsahar H, Gravier C, Laforest F. Zero-shot question generation from knowledge graphs for unseen predicates and entity types[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018: 218-228.
- [15] Wang H, Zhang X, Wang H. A neural question generation system based on knowledge base[C]//Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, 2018:133-142.
- [16] Gu J, Lu Z, Li H, et al. Incorporating copying mechanism in sequence-to-sequence learning[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 1631-1640.
- [17] Cai D, Lam W. Graph transformer for graph-to-sequence learning [J]. arXiv preprint arXiv: 1911.07470, 2019.
- [18] Koncel-Kedziorski R, Bekal D, Luan Y, et al. Text generation from knowledge graphs with graph transformers[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 2284-2293.
- [19] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.
- [20] 胡月. 面向大规模知识库的开放域问题生成技术研究[D].武汉:华中师范大学硕士学位论文,2020.
- [21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of Advances in Neural Information Processing Systems, 2017: 5998-6008.
- [22] Bollacker K, Cook R G, Tufts P, et al. FreeBase: A shared database of structured general human knowledge[C]//Proceedings of National Conference on Artificial Intelligence, 2007: 1962-1963.

- [23] Bordes A, Weston J, Collobert R, et al. Learning structured embeddings of knowledge bases[C]// Proceedings of the 25th Conference on Artificial Intelligence, 2011: 301-306.
- [24] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes[C]// Proceedings of the 28th Conference on Artificial Intelligence, 2014: 1112-1119.
- [25] Lin Hailun, Liu Yong, Wang Weiping, et al. Learning entity and relation embeddings for knowledge resolution[J]. Procedia Computer Science, 2017, 108: 345-354.
- [26] Lin Y, Liu Z, Luan H, et al. Modeling relation paths for representation learning of knowledge bases[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015: 705-714.
- [27] Huang X, Zhang J, Li D, et al. Knowledge graph embedding based question answering[C]// Proceedings of the 12th ACM International Conference on Web Search and Data Mining, 2019: 105-113.
- [28] Papineni K, Roukos S, Ward T, et al. BLEU: A method for automatic evaluation of machine translation[C]// Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002: 311-318.
- [29] Denkowski M, Lavie A. Meteor universal: Language specific translation evaluation for any target language [C]// Proceedings of the 19th Workshop on Statistical Machine Translation, 2014: 376-380.
- [30] Lin C Y. ROUGE: A package for automatic evaluation of summaries[C]// Proceedings of the Workshop on Text Summarization Branches Out at ACL, 2004.
- [31] Pennington J, Socher R, Manning C D. GloVe: Global vectors for word representation[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2014: 1532-1543.
- [32] Abadi M, Barham P, Chen J, et al. Tensorflow: A system for large-scale machine learning [C]// Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, 2016: 265-283.
- [33] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv: 1412.6980, 2014.



胡月(1994—), 硕士, 主要研究领域为自然语言处理。

E-mail: huyue2017@mails.ccnu.edu.cn



周光有(1983—), 通信作者, 博士, 教授, 主要研究领域为自然语言处理、信息抽取和问答系统。

E-mail: gyzhou@mail.ccnu.edu.cn