

基于Graph Transformer的知识库问题生成

胡月

华中师范大学
计算机学院
湖北 武汉

huyue2017@mails.ccnu.edu.cn

周光有

华中师范大学
计算机学院
湖北 武汉

gyzhou@mail.ccnu.edu.cn

摘要

知识库问答依靠知识库推断答案需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，该任务是利用知识库三元组自动生成问题。现有方法仅由一个三元组生成的问题简短且缺乏多样性。为生成信息量丰富且多样化的问题，本文采用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上的实验结果证明了该方法有效性。

关键词： 问题生成；知识库；语义表征；知识库问答

Question Generation from Knowledge Base with Graph Transformer

Yue Hu

Central China Normal University
School of Computer
Wu Han Hu Bei
huyue2017@mails.ccnu.edu.cn

Guangyou Zhou

Central China Normal University
School of Computer
Wu Han Hu Bei
gyzhou@mail.ccnu.edu.cn

Abstract

Knowledge base question answering requires a large number of question answering pairs when relying on the knowledge base to infer answers. However, building a large-scale and accurate data set is not only expensive, but also limited by factors such as domain. To alleviate the problem of data labeling, 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. However, existing methods only use a triple to generate questions that are short and lack diversity. To generate questions with rich and diverse information, this paper use 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

1 引言

问答(Question Answer, 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对作为标记数据, 例如基于知识库的SimpleQuestions(Bordes et al., 2015)数据集, 基于文本的Wiki QA(Vrandečić and Kröttsch, 2014)数据集, 以及阅读理解所用的SQuAD(Rajpurkar et al., 2016)数据集等。这些数据集受大小等限制, 且标注花费昂贵。并且随着问答系统、阅读理解等领域的发展, 对标注数据的需求愈加强烈, 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任务开始逐渐作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。随着深度学习研究的推入, 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。并且, 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(2013)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(2015)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标签决定。Seyler等人(2017)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 (Seq2Seq) 框架(Sutskever et al., 2014)和编码-解码结构(Cho et al., 2014)为自然语言生成(张建华和陈家骏, 2006)任务带来了新的研究方向, 并在机器翻译(Luong et al., 2015)、智能对话(贾熹滨等人, 2017)等任务获得优异的成果, 一系列以基于神经网络的QG系统被提出。Serban等人(2016)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30M的标签数据。Khapra等人(2017)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(2018)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(2018)在Seq2Seq基础上加入复制机制(Gu et al., 2016), 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(2019)和Koncel-Kedziorski(2019)等人工作的启发, 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(Devlin et al., 2019)加强的词级编码层。本文预先将知识库中实体、关系构成知识图, 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(Vaswani et al., 2017)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)(Cho et al., 2014)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SimpleQuestions上进行了实验, 实验的评测结果表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, 本文率先提出使用基于Graph Transformer的方法为实体获取丰富的背景信息。(2) 本文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且表达形式多样化。

2 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(2019)和Cai(2019)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, 结合了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。

2.1 模型结构

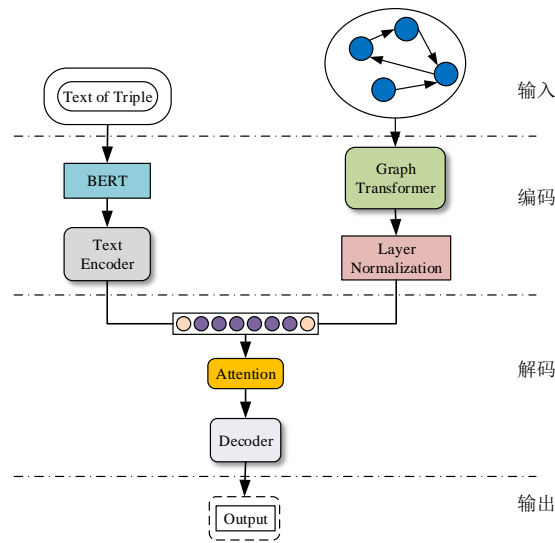


图 1. 模型结构图，双编码层在解码端连接

本节将介绍基于Graph Transformer(GT)的知识库问题生成模型(简称GT-KBQG)。GT-KBQG模型的结构如图1所示，由三个模块组成，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 $e_{triple} = (e_s, p_r, e_o)$ 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进行结合作为解码器的隐藏层初始化状态，解码器结合Attention机制(Luong et al., 2015)获得上下文向量，以此加强问题的生成效果。

2.2 基于Graph Transformer的图编码层

为了让实体之间更加紧密和结构化，有学者将其构成图的形式(实体为结点，关系为边)，即知识图谱。本文利用已构成的知识图赋予数据集中的三元组丰富的背景网，结合Transformer结构，在丰富的知识背景网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，单个实体的语义向量也更加丰富。

2.2.1 知识构图预处理

英文知识库FreeBase(Bollacker et al., 2007)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，故其原始实体也是以id形式，为了模型能够获取更多的实体背景信息，本文将大规模知识库预先训练获得的知识图作为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嵌入(Knowledge Graph Embedding, KG-Emb)(Bordes et al., 2011)。

KG-Emb是为了将高维的知识图表示为低维谓词、实体的表示(P 和 E)。为实现这一目标，Bordes等人(2011)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(2014)提出的TransH处理一对多或者多对一关系，与TransE直接测量 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(2017)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等(Lin et al., 2015)多种图嵌入表示方法。

本文选用TransE方法对三元组事实进行图的预处理，将实体作为整体转为向量的表达。选择TransE主要有以下几个考虑：首先，与它的改善结构如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等人(2019)在FB2M数据集上所提供的基于TransE的KG-Emb。

2.2.2 Graph Transformer

本文使用预处理后的知识图作为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(2017)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，解决了RNN的在长序列上的顺序计算结构的缺点。

GT作为图编码器，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，如图2表示的为单个网络块，左边部分的 e_o 、 e_s 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。

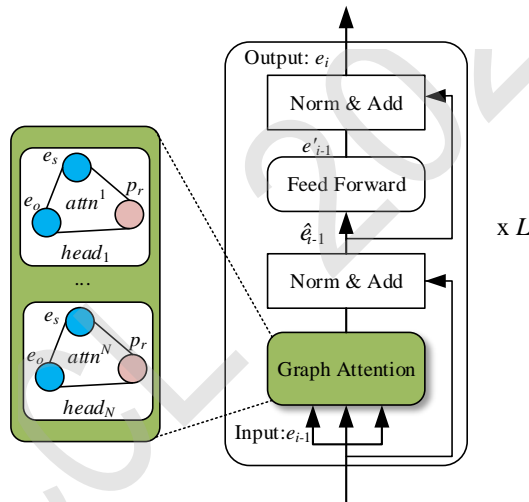


图 2. Graph Transformer结构

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的具体计算如公式(1)所示：

$$\hat{e}_i = e_i + \parallel_{j=1}^N attn^j(q_i, k_i)v_i \quad (1)$$

在公式(1)中，其中 \parallel 表示 N 个 $attn$ 的连接操作($j \in N$)， $attn^j$ 为点积计算，如公式(2)所示，为了降低点积造成的阻碍梯度流趋势，参照Vaswani等人(2017)的方法，将结果缩小 $\sqrt{d_k}$ 之后再利用 $softmax$ 函数进行归一化。

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

公式(1)和公式(2)中出现的 q_i ， k_i ， v_i 是第 i ($i < L$)个堆叠块对输入进行线性变换后的 d_k 维的向量表示，如公式(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 $i - 1$ 层第二个LN层标准化计算得来，即为第 i 层的输入,其中LN为标准化层。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，如公式(4)和(5)所示:

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

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

如公式(5)所示，第 i 层的最终输出作为第 $i + 1 (\leq L)$ 层的输入。公式(4)中， $FF(\cdot)$ 为两层前馈网络，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，如公式(6)所示。

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

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

$$e_N = LN_{output}(e_N) \quad (7)$$

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。

2.3 基于BERT增强词级表示的编码层

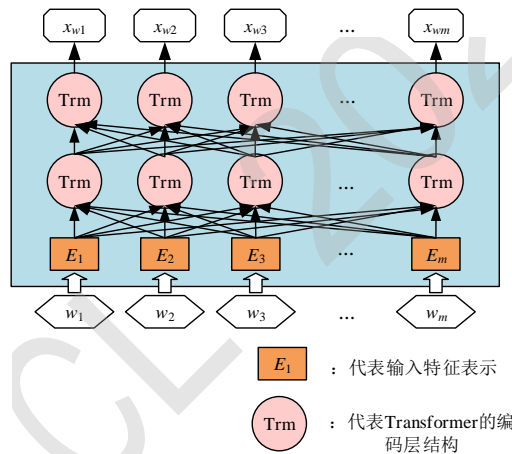


图 3. BERT预处理结构图

在2018年，Jacob Devlin等人(Devlin et al., 2019)提出BERT预训练模型在NLP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上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，再自行将所需数据放入就可以直接使用。SimpleQuestions数据集中实体原始为独立的id形式，而在目标问题中为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，所以本节也将三元组的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有效。考虑到BERT在编码方面的优异成绩，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，实体对应于科学术语，由多个词来表达的，三元组的词组输入为短字符串，因此我们采用双向GRU进行编码来嵌入三元组的词语粒度表示。我们通过BERT预处理泛化每个词的嵌入向量的语义信息，再采用

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

双向GRU的最后一个隐藏状态作为每个词生成d维嵌入，该状态位于实体词组中每个词的嵌入之上的。

如图3所示，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 $x_w = \{x_{w_1}, \dots, x_{w_i}, \dots, x_{w_m}\}$ 。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1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。公式(8)和(9)可表示该过程。

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

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

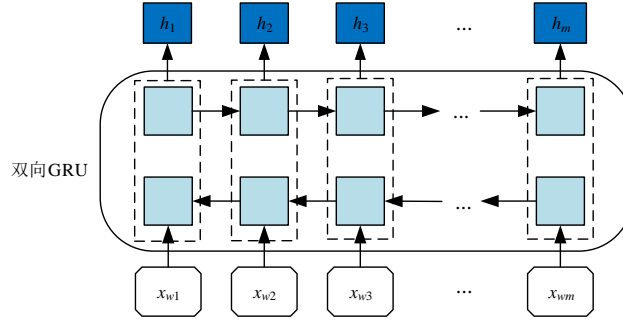


图 4. Text Encoder结构图

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 $h_i = [\vec{h}_i; \overleftarrow{h}_i]$ 。

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 经过公式(10)成为解码层的隐藏状态的初始化值。

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

2.4 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，发现由于优良的并行结构使得模型计算时存在注意力过长，导致了生成的序列中出现与对应输入不相关的词汇。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 个单词的概率分布由公式(11)所示。

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

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，如公式(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 的详细计算由公式(13)和公式(14)给出。

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

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

其中， α_t 为第 t 时间步的对齐向量，详细计算参照公式(15)。

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

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。

3 实验结果与分析

3.1 实验数据

本文使用了具有id形式的英文数据集SimpleQuestions，表1所示为文本与id形式的问答对示例。本文的输入分为两部分，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中所有的事实都是基于FreeBase的。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(2019)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。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。

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/0mjn2 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?

表 1. SimpleQuestions数据集id和文本形式样例

#	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

表 2. 数据集统计信息

3.2 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(Papineni et al., 2002)，METEOR(Banerjee and Lavie, 2005)，ROUGE-L(Lin, 2004)。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在翻译任务中最先被使用，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考虑的是单个单词的匹配，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,其目的是解决BLEU标准固有的一些缺点。它还考虑到了一些其它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。

目标问句	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

表 3. BLEU值得分举例,表中BLEU分数值为BLEU-1至BLUE-4的得分均值

人工评估(Human Evaluation): 尽管自动评测在某种程度来说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上不同时,许多隐藏着的限制就显示出来了。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的说明这个问题,如表3。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。此时,人工来评测给出句子的流畅度与相关度是必不可少的。

3.3 基准模型

本文做了多个对比实验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(Vaswani et al., 2017): Transformer是当前在机器翻译任务上非常火热的模型。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(ElSahar et al., 2018): 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(Serban et al., 2016): 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 et al., 2018): Wang等人(2018)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。

3.4 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(Pennington et al., 2014)预训练好词嵌入对词向量初始化。本文基于Tensorflow平台(Abadi et al., 2016),采用Adam优化算法(Kingma and Ba, 2014)。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。

3.5 结果分析

(一) 自动评测

如表4所示,GT-KBQG在BLEU-1、BLEU-2、Meteor和ROUGE-L指标上提升效果明显。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

#	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

表 4. 自动评估结果

比, GT-KBQG在Meteor (+2.98)和ROUGE-L (+9.96)指标上提升效果非常明显, 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。CopyNet模型对比其他的基准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BLUE-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。

(一) 人工评价

#	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%

表 5. 人工评估结果

由于语言表述的复杂性, 自动评测只能表现出针对目标问题这一单一的参考文本的效果, 无法完全准确反映生成的问题是否合适, 以及对生成问句与对应三元组的信息相关性、背景信息的描述等方面有所欠缺。因此, 本文采用人工评价的方式对GT-KBQG模型生成问句的多样性进行评估, 表5中Sim.表示similarity, 该行表示与参考问句几乎完全相似(可能冠词不一致)的句子占有所有被选问句的比例, 表中的模型所占比例都在一半左右, 具有较高的一致性。本文除了要考虑生成问句的一致性外, 更加要关注是否具有多样性, 对背景信息是否丰富进行标注, 即表5中的Var.(variety)。

本文在测试集中随机抽取三份数据用于人工评价, 每份有三百个三元组-问题对, 同时每份以三个人参与, 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%)的结果都低, 可以推测出目标问题较多比较简单, 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, 从而生成的问句中的实体的表述信息更多。

(三) 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选取目标问题和Enc-Dec模型的样例与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生成出更加具有描述性的问题。如示例1和示例3中Enc-Dec所生成的问题正确但非常简洁, 对于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

#	Facts	Models	Generated Questions
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?

表 6. 样例对比

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，尽管在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，BLEU等自动指标效果也会愈低。

4 总结与展望

本文提出了基于Graph Transformer的问题生成模型，缓解模型生成单一化的问题。与额外加入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，但依旧存在许多问题待解决，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，需要重新构造一个更完善的图，在后续工作上可以考虑自行构造出更大更完善的知识图。由于时间环境、设备等因素的限制，本文提出的GT-KBQG模型只基于SimpleQuestions上实验验证，接下来将会在多种数据集上进行研究，以及将进一步深入消融实验分析。

致谢

本文的工作作为硕士毕业论文的一部分，受到国家自然科学基金（No. 61972173）支持。感谢匿名审稿人对我们工作提出的建设性修改意见。

参考文献

- Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek Gordon Murray, Benoit Steiner, Paul A. Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu and Xiaoqiang Zheng. 2016. TensorFlow: A System for Large-Scale Machine Learning. In *12th USENIX Symposium on OSDI*, page 265–283.
- Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72.
- Kurt D. Bollacker, Robert P. Cook and Patrick Tufts. 2007. Freebase: A Shared Database of Structured General Human Knowledge. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, Vancouver, British Columbia, Canada*, pages 1962–1963.
- Antoine Bordes, Jason Weston, Ronan Collobert and Yoshua Bengio. 2011. Learning Structured Embeddings of Knowledge Bases. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, California, USA*.
- Antoine Bordes, Nicolas Usunier, Sumit Chopra and Jason Weston. 2015. Large-scale Simple Question Answering with Memory Networks. <http://arxiv.org/abs/1506.02075>.
- Deng Cai and Wai Lam. 2019. Graph Transformer for Graph-to-Sequence Learning. <http://arxiv.org/abs/1911.07470>.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *EMNLP*, pages 1724–1734.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT, Minneapolis, MN, USA, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Daniel Duma and Ewan Klein. 2013. Generating Natural Language from Linked Data: Unsupervised template extraction. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*, pages 83–94.
- Hady ElSahar, Christophe Gravier and Frédérique Laforest. 2018. Zero-Shot Question Generation from Knowledge Graphs for Unseen Predicates and Entity Types. In *NAACL-HLT, New Orleans, Louisiana, USA, Volume 1 (Long Papers)*, pages 218–228.
- Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O. K. Li. 2016. Incorporating Copying Mechanism in Sequence-to-Sequence Learning. In *ACL (Volume 1: Long Papers)*, pages 1631–1640.
- Xiao Huang, Jingyuan Zhang, Dingcheng Li and Ping Li. 2019. Knowledge Graph Embedding Based Question Answering. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, Melbourne, VIC, Australia*, pages 105–113.
- 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, 43(009):1344-1356.
- Mitesh M. Khapra, Dinesh Raghu, Sachindra Joshi and Sathish Reddy. 2017. Generating Natural Language Question-Answer Pairs from a Knowledge Graph Using a RNN Based Question Generation Model. In *EACL, Valencia, Spain, Volume 1: Long Papers*, pages 376–385.
- Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. In <http://arxiv.org/abs/1412.6980>.
- Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata and Hannaneh Hajishirzi. 2019. Text Generation from Knowledge Graphs with Graph Transformers. In *NAACL-HLT, Minneapolis, MN, USA, Volume 1 (Long and Short Papers)*, pages 2284–2293.
- Yankai Lin, Zhiyuan Liu, Huan-Bo Luan, Maosong Sun, Siwei Rao and Song Liu. 2015. Modeling Relation Paths for Representation Learning of Knowledge Bases. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP, Lisbon, Portugal*, pages 705–714.

- Hailun Lin, Yong Liu, Weiping Wang, Yinliang Yue and Zheng Lin. 2017. Learning Entity and Relation Embeddings for Knowledge Resolution. In *International Conference on Computational Science, ICCS, Zurich, Switzerland, Volume 108*, pages 345–354.
- Chin Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of summaries. In *In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*.
- Thang Luong, Hieu Pham and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP, Lisbon, Portugal*, pages 1412–1421.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *ACL, Philadelphia, Pennsylvania, USA*, pages 311–318.
- Jeffrey Pennington, Richard Socher and Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, Doha, Qatar*, pages 1532–1543.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev and Percy Liang. 2016. SQuAD: 100, 000+ Questions for Machine Comprehension of Text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP, Austin, Texas, USA*, pages 2383–2392.
- Iulian Vlad Serban, Alberto García-Durán, Çağlar Gülçehre, Sungjin Ahn, Sarath Chandar, Aaron C. Courville and Yoshua Bengio. 2016. Generating Factoid Questions With Recurrent Neural Networks: The 30M Factoid Question-Answer Corpus. In *ACL, Germany, Volume 1: Long Papers*, pages 588–598.
- Dominic Seyler, Mohamed Yahya and Klaus Berberich. 2015. Generating Quiz Questions from Knowledge Graphs. In *Proceedings of the 24th International Conference on World Wide Web, Florence, Italy*, pages 113–114.
- Dominic Seyler, Mohamed Yahya and Klaus Berberich. 2017. Knowledge Questions from Knowledge Graphs. In *ICTIR, Amsterdam, The Netherlands*, pages 11–18.
- Ilya Sutskever, Oriol Vinyals and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In *NeurIPS*, pages 3104–3112.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA*, pages 5998–6008.
- Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. In *Commun. ACM, Volume 57, 57(10)*:78–85.
- Zhen Wang, Jianwen Zhang, Jianlin Feng and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, Québec, Canada*, pages 1112–1119.
- Hao Wang, Xiaodong Zhang and Houfeng Wang. 2018. A Neural Question Generation System Based on Knowledge Base. In *NLPCC*, pages 133–142.
- 张建华, 陈家骏. 自然语言生成综述[J]. 计算机应用研究, 2006(08):1-3.