#	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 Merrienboer, 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, Çaglar 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 Vrandecic 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 *roceedings 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.