

文章编号:

基于 Transformer 的 AMR-to-Text 生成*

朱杰¹, 李军辉¹, 钱龙华¹

(1. 计算机科学与技术学院, 苏州大学, 江苏省苏州市 215006)

摘要: 抽象语义表示到文本 (AMR-to-text) 生成的任务是给定 AMR 图, 生成相同意义表示的文本。可以把此任务当成一个从源端 AMR 图到目标端句子的翻译任务。然而, 传统的序列到序列 (seq2seq, 简称 S2S) 的方法, 使用循环递归网络进行编码, 并不能很好的解决长短期依赖的问题。当前最好的性能是图到序列 (graph2seq, 简称 G2S) 的模型, 使用了图模型直接对 AMR 图结构进行编码, 但对于非直接相连的节点依然会损失大量的结构信息。针对上述问题, 基于 seq2seq 框架, 提出了一种直接而有效的 AMR-to-text 生成方法。在这项工作中, 引入了当前最优的 seq2seq 模型 Transformer 作为基准模型, 并且使用字节对编码 (简称 BPE) 和共享词表的方法联合来对未登录词 (简称 OOV) 的问题进行处理。在现有的两份英文标准数据集上, 实验结果都取得了显著的提升, 达到了新的最高性能。

关键词: AMR-to-text 生成; 序列到序列模型; 字节对编码; 共享词表; 未登录词;

中图分类号: TP391

文献标识码: A

AMR-to-Text generation based on Transformer

Jie Zhu¹, Junhui Li¹, Longhua Qian¹

(1. School of Computer Science and Technology, Soochow University, Suzhou, 215006)

Abstract: Abstract semantic representation to text (AMR-to-text) generation is the task of generating text of the same meaning for a given AMR graph. This task can be viewed as a translation task from the source AMR graph to the target sentence. However, traditional sequence-to-sequence (seq2seq) models use a recursive neural network for encoding, but do not solve the problem of long-term and short-term dependencies well. The current best performance is a graph-to-sequence (graph2seq) model that uses a graph model to directly encode the AMR graph structure. Although the problem of non-local semantic information loss can be solved to a large extent, the model is relatively complicated. Aiming at the above problems, based on the seq2seq framework, a direct and effective AMR-to-text generation method is proposed. In this work, the current optimal seq2seq model Transformer is introduced as the benchmark model, and the problem of out of vocabulary are processed using a combination of byte pair coding (BPE) and shared vocabulary. On the two English benchmark datasets, the experimental results show that our performance is significantly improved, and reaches a new state-of-the-art.

Key words: AMR-to-text generation; Sequence to sequence model; BPE; shared vocabulary; out of vocabulary;

* 收稿日期: 定稿日期:

基金项目: 自然科学基金 (61876120);

作者简介: 朱杰 (1996—), 男, 硕士研究生, 主要研究方向语义分析; 李军辉 (1983—), 男, 副教授, 主要研究方向自然语言处理; 钱龙华 (1966—), 男, 教授, 主要研究方向自然语言处理。

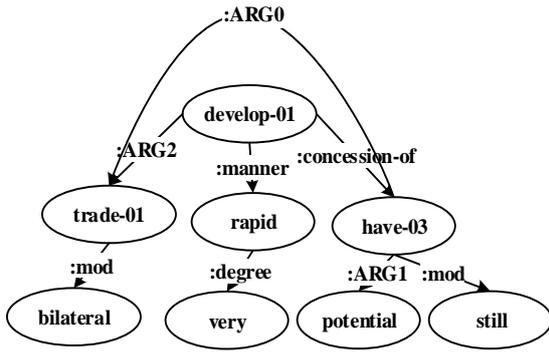


图1 “Bilateral trade has developed very rapidly but still had potential.” 表示成 AMR 图的一个例子

0 引言

抽象语义表示 (Abstract Meaning Representation, 简称 AMR) 是一种新型的语义表示方法^[1-8], 它将自然语言文本以句子为单位抽象成单根有向无环图^[9]。如图 1 所示, “develop-01” 和 “have-03” 等图结点称做为概念 (concept), 自然语言中的实词被映射为 AMR 图中的概念结点, 结点之间的边表示两个概念之间的语义关系, 如 “:ARG0” 和 “:mod” 等。AMR 已经被广泛应用于机器翻译^[10], 问答系统^[11], 事件抽取^[12]等自然语言处理相关任务中。与此同时, AMR-to-text 生成的任务也越来越受到人们的关注。

AMR-to-text 生成是给定 AMR 图, 自动生成语义上一致的自然语言文本。AMR-to-text 生成至少存在着下面几方面挑战: (1) AMR 图中的并没有包含时态等信息; (2) AMR 图中并不包含介词等功能词; (3) AMR 图是一个有向无环图, 但自然语言句子却是一个线性序列, 因此, 在 AMR 图中相近的两个概念结点, 生成的文本中对应的片段并不一定相邻。

最初的工作都是采用基于统计^[1-3]的方法, 之后 Konstas 等^[4]人将该任务看作机器翻译问题, 并成功地将基于深度神经网络的序列到序列 (seq2seq) 模型应用于该任务上。seq2seq 模型的输入是一个线性化单元序列。因此, 针对本任务来说, 需要对源端

AMR 图进行线性化处理。虽然循环神经网络可以在一定程度上解决长短期依赖的问题, 但当序列过长时, 其并不能很好的获取到非局部信息, 因此, 线性化 AMR 图就会导致损失许多结构化信息, 特别是当 AMR 图中结点较多时。比如, 位于同一个父亲结点下的兄弟结点, 在经过线性化后, 可能会相距很远, 循环神经网络很难获取到它们之间的结构化信息^[4]。因此, 为了更好的获取到非局部信息, Beck 等^[5], Song 等^[6], Damonte 等^[7], Guo 等^[8]提出了图到序列 (graph2seq) 的框架, 使用图模型直接对 AMR 图进行编码。虽然相比于循环递归网络可以获取到更多的非局部信息, 但是依然会损失图中非直接相连节点之间的结构信息。

需要指出的是, 上面相关工作的图模型大都是基于 LSTM 网络的, 即在传统的 LSTM 网络中, 利用 AMR 图的边信息, 调整上一个输入单元的隐藏状态, 使得模型能够对 AMR 图进行建模, 捕获图结构信息。相比于 LSTM 模型, 目前最新提出的基于自注意力机制的 seq2seq 模型 Transformer^[13]能够更有效地捕获全局信息, 更好地弥补 LSTM 模型中存在的对长输入序列学习效果欠佳问题。因此, 本文将探索基于 Transformer 的 AMR-to-text 生成。特别地, 鉴于 AMR-to-text 任务本身的特点, 即源端和目标端存在着大量相同词汇, 本文采用了源目标共享词汇表的做法; 鉴于数据稀疏问题, 本文采用了字节对编码 (BPE) 机制^[14]。此外, 本文也参照 Konstas 等^[4]人提出的“对训练”的方法, 探索了外部未标注语料对本文模型性能的影响。

基于现有的两份英文 AMR 数据集 (LDC2015E86 和 LDC2017T10) 上的实验结果表明, 本文的模型的性能超过了目前已有的模型, 说明了 Transformer 模型能够在很大程度上学习得到隐式的深层的源端结构化信息。使用外部未标注语料进一步说明了本文模型的优越性。

1 相关工作

目前 AMR-to-text 生成的方法可以大致分为三类：基于统计的方法、seq2seq 方法以及 graph2seq 方法。

早期的研究均是采用基于统计的方法。其中，Flanigan 等^[3]设计出一个基于对齐的树到字符串（tree-to-string）的转换规则。将输入的 AMR 图根据重入（re-entrances）节点（具有两个及两个以上父节点）划分成树结构，然后使用规则将其翻译为句子。Song 等^[2]使用同步节点替换语法同时进行 AMR 图的解析与生成。Pourdamghani 等^[1]将输入的 AMR 图根据广度优先遍历进行线性化，然后使用基于短语的机器翻译系统去生成。

随后，Sutskever 等^[15]表明了深度神经网络的优越性。Konstas 等^[4]人将 seq2seq 框架应用在这项任务上，并且取得了当时最好的性能。此外，他们也提出了一种“对训练”的方法，从外部未标注语料 Gigaword 中抽取在模型词表中而又不与训练和测试集重合的句子，然后使用 AMR 分析模型（AMR parser）将句子解析成 AMR 图，得到大量外部有标签的语料，之后再联合 AMR 生成器进行联合训练。为了限制生成的文本具有更合理的句法，Cao 和 Clark^[16]先使用句法模型生成最优目标端句法结构，然后再利用句法信息辅助词法模型生成句子。

为了减少 AMR 图线性化之后的信息损失，Song 等^[6]提出了一种新颖的 graph2seq 模型，使用图状态 LSTM 通过每步的迭代交换相邻节点的信息来更新节点。同时也对每一个节点增加了一个向量单元用来保存历史信息。Beck 等^[5]引入了一个基于门控图神经网络（GGNNs）的编码器，将 AMR 图结构完整的融入模型当中，将边信息也转化为节点，解决了参数爆炸问题的同时也给了解码器更多的信息。着重于解决重入节点问题。Damonte 和 Cohen^[7]提出了一种堆栈式编码器，由图卷积神经网络和双向 LSTM 组合而成。Guo 等^[8]人提出了一种深度连接图卷积网络（GCN）更好的获取局部和非局部

信息。

但是，以上提及的图模型大部分是基于 LSTM 神经网络，即使通过修改 LSTM 网络的输入单元用于对图结构信息建模，仍然会导致损失大量非直接相连结点之间的结构信息。

2 基于 Transformer 的 AMR-to-Text 生成

本文使用当前性能最佳的 seq2seq 模型 Transformer^[13]作为基准模型，该模型最初被应用于机器翻译和句法解析的任务当中。相关研究表明它的自注意力机制（self-attention）可以较好地捕获源句的全局结构化信息。

为了使 Transformer 模型适用于 AMR-to-text 生成的任务，需要预先对 AMR 图进行线性化处理。针对数据稀疏性的问题，本文采用了字节对编码（BPE）的机制。同时考虑到该任务自身的特点，源端和目标端存在大量相同的词汇，本文采用了源目标端共享词汇表的做法。此外，本文也遵循着 Konstas 等^[4]人提出的“对训练”的方法，研究了加入外部未标注语料对于模型性能的影响。

2.1 序列到序列模型

本文将 Transformer 模型作为 AMR-to-text 生成的基准模型。Transformer^[13]抛弃了传统的卷积神经网络^[17]（CNN）和循环神经网络^[18]（RNN）顺序的结构，整个网络完全由自注意力机制组成。其编码器由多个相同的层堆栈而成，每层有两个子层，一个是多头自注意力机制，另一个是前馈神经网络。解码器部分也是由多个层堆栈而成。与编码器的唯一区别就是每层比编码器多了一个对编码器堆栈的最后一层输出进行计算的基于上下文的多头自注意力机制。详细内容可以参阅 Vaswani 等^[13]论文。

传统用在该任务上的 seq2seq 模型，双向 LSTM，它只能从左到右依次计算或者从右到左依次计算，这种机制带来了两个问



图2 一个 AMR 图线性化示例

题: 1. 时间片 t 的计算依赖于 $t-1$ 时刻的计算结果, 这样大大限制了模型的并行能力; 2. 顺序计算的过程中信息的丢失。这种序列性计算的机制意味着上下文之间的语义关联会随着状态的传递不断地减弱, 对长文本不友好, 而且计算是线性的, 通常需要一定的时间^[19]。特别是在 AMR-to-text 任务当中, AMR 图线性化之后的长度通常过长, 这就会导致图结构信息的大量丢失。

Transformer 模型很好地解决了上述的两个问题, 首先使用了自注意机制, 将序列中的任意两个位置之间的距离缩小到一个常量, 可以直接获取到全局信息。其次, 它不是类 RNN 之类的顺序结构, 具有很好的并行性, 而且运算机制都是矩阵的运算, 因此特别符合现有的 GPU 框架, 有着模型复杂度低, 效率高的优点。在机器翻译等任务上的实验结果表明, Transformer 都要优于循环递归网络的 seq2seq 模型。

2.1 线性化预处理

本文综合了 Barzdins^[20]、Noord^[21]以及 Song^[6]等实验中的预处理方法。该方法首先通过删除变量和 wiki 链接获得简化的 AMR 图, 即 AMR 树。因为 AMR 图中的变量标

签仅表示共同引用一个节点, 不具有任何有意义的信息, 所以也将其删除。最后, 将 AMR 图中每行的换行符替换为空格符, 从而就完成了 AMR 图线性化的转换。图 2 给出了一个 AMR 图线性化的示例。值得注意的是, 本文在对图进行线性化操作的时候, 对于重入节点的处理是将其映射为两种不同的词, 即在目标位置再复制一个重入节点。最后, 将括号前后加入空格符, 因为括号要当作独立的标签输入模型。

对于目标端的处理, 我们使用斯坦福¹的词汇化工具对句子进行了词汇化处理。

2.2 BPE 方法

低频词的翻译一直是神经机器翻译的主要难点之一。为了解决这个难点, Konstas 等^[4]使用了匿名化策略, 但是需要人为的去制定规则, 耗时耗力, 成本太高。Song 等^[6]使用了 charLSTM 加 copy 机制, 但是需要模型自动去学习什么时候从源端 copy 词汇, 增加了模型的复杂度。受机器翻译领域解决低频词翻译方法的启发, 本文采用一种直接而有效的可行方案 ---- 字节对编码 (Byte Pair Encoding, 简称 BPE) 方法^[14,22]。

¹ <https://nlp.stanford.edu/software/tokenizer.html>

BPE 的基本思想是将原词汇单元拆分成更小粒度的高频子字单元。对于原本不在词表中的单词，seq2seq 模型一般会将它们标记为<UNK>。BPE 方法就是将未在词表中出现的未登录词拆解成常用的子字。通过翻译子字部分将原来被标识为<UNK>的单词翻译了出来，从而极大地提高了翻译质量。并且只需要在翻译前使用脚本对语料进行 BPE 处理，翻译后去掉“@@”标识符就可以完成，不需要对 seq2seq 模型增加任何参数。恰恰相反的是，3.3 节的对比实验表明，通过 BPE 之后的词表会更小，这也就让模型的参数更少，训练更有效率。BPE 方法对语料处理的具体效果通过如下例子来说明：

例(a) it begins and ends with romneycare .

例(b) it begins and ends with romne@@ y@@ care .

假如例(a)是训练语料中的一个句子，“romneycare”在训练语料中出现的次数较少，就会导致词汇表中没有包含它。经过 BPE 处理之后如例(b)所示，将句子中的低频词“romneycare”拆分成了 3 个子单元 romne@@、y@@和 care。

2.3 使用无标签的外部语料

获取一份 AMR 图和句子配对的平行语料是十分昂贵的，现存的最大 AMR 语料也仅仅只有 3 万多条。神经网络模型也一直遭受着数据稀疏性的问题。为了解决这个问题，本文遵循了 Konstas 等^[4]人提出的“对训练”方法进行了外部语料实验。本文从 Gigaword 中抽取大量无标签的句子，并且这些句子中的词汇是包含在训练集的词汇表中。然后使用一个预训练完成的 AMR 解析器生成与句子配对的伪 AMR 图，这样就组成了伪平行语料。得到伪平行语料之后，将其放入本文的模型中进行训练。训练的基本方法是在每一轮（epoch）伪语料迭代完成之后，再使用人工标注语料进行调优（fine-tuning）。

3 实验

3.1 数据集

本文使用现存的两份英文标准 AMR 语料集（LDC2017T10 和 LDC2015E86）作为实验的数据集，分别包含了 36521 个和 16833 个训练样例，并且它们共用了 1368 个验证样例以及 1371 个测试样例。每个样例都包含了一个 AMR 图和其对应的句子。本文分别使用 BPE 操作数为 10K 和 20K 对两份语料进行子词的划分。

融合外部语料的实验，本文从 Gigaword 中抽取句子，使用的 AMR 解析器是 Flanigan 等^[23]人提出的 JAMR 解析器。需要注意的是，对于外部语料，本文同样也做了 BPE 的操作。

3.2 实验设置

本文从 BPE 处理之后的训练集中，根据词频抽取单词形成词汇表，参考 Ge 等^[24]的工作，共享了源目标端的词汇表。词向量使用的是随机初始化的嵌入（embedding）。

本文使用 OpenNMT 的框架作为 Transformer 的基准模型。在超参的设置上，模型的编码器和解码器层数默认为 6，多头自注意机制（Multi-head attention）的头数设置为 8，隐藏层大小和维度设置为 512，批处理大小为 4096，词汇表大小为 20K，批处理以 token 为单位。同时，模型使用学习率为 0.5，beta 1=0.9，beta 2=0.998 的 Adam 优化算法^[25]，将交叉熵作为损失函数。使用束搜索宽度（beam size）为 5 进行解码。其余参数使用该系统的默认值。在本文所有实验中，模型在 K40 GPU 上训练 30K 步，报告了从验证集中挑选出的单个模型最高的结果。

为了更好的说明模型的性能，本文采用了 BLEU^[26]，Meteor^[27]和 CHRF++^[28]三种评测工具。其中 BLEU 是基于语料级的评测，而后两者是基于句子级的评测。后两者的评测分数相对来说更接近于人工评测。

表 1 测试集实验结果和现存其它模型的比较 († 代表是图模型, * 代表是 seq2seq 模型)

AMR-to-text 生成模型	数据集	BLEU	Meteor	CHRF++
Konstas 等 ^[4] 人 (2017) *		22.00	-	-
Cao 和 Clark 等 ^[16] 人 (2019) *		23.5	-	-
Song 等 ^[6] 人 (2018) †		23.30	-	-
Damonte 和 Cohen 等 ^[7] 人 (2019) †	LDC2015E86	24.40	23.60	-
Guo 等 ^[8] 人 (2019) †		25.7	-	-
本文		25.50	33.16	59.88
<hr/>				
Cao 和 Clark 等 ^[16] 人 (2019) *		26.8	-	-
Beck 等 ^[5] 人 (2018) †		23.3	-	50.4
Damonte 和 Cohen 等 ^[7] 人 (2019) †	LDC2017T10	24.54	24.07	-
Guo 等 ^[8] 人 (2019) †		27.6	-	57.3
本文		27.43	34.62	61.85
<hr/>				
Konstas 等 ^[4] 人 (2017) (200K) *		27.4	-	-
Song 等 ^[6] 人 (2018) (200K) †		28.2	-	-
Guo 等 ^[8] 人 (2019) (200K) †		31.6	-	-
Konstas 等 ^[4] 人 (2017) (2M) *	LDC2015E86 + 外部语料	32.3	-	-
Song 等 ^[6] 人 (2018) (2M) †		33.0	-	-
Konstas 等 ^[4] 人 (2017) (20M) *		33.8	-	-
本文 (200K)		37.45	-	-

3.3 对比实验

表 2 LDC2015E86 验证集上的对比实验结果

模型	BLEU	Meteor	CHRF++	# P
基准模型	18.77	28.04	51.88	61.7M
+BPE	23.24	31.78	58.43	56.3M
+共享词表	23.02	31.60	58.09	52.6M
+两者	24.93	33.2	60.3	49.2M

如上文所述,本文使用了 BPE 方法^[14,22]来处理低频词的翻译问题,并且采用了共享词表的方法来进行训练。为了更好的了解到 BPE 机制和共享词表对模型性能贡献度,本文在 LDC2015E86 验证集上做出了如下的对比实验。表 2 给出了该实验的结果。

从表 2 中可以看到,通过加入 BPE 机制和共享词表的操作之后,模型的性能取得了显著提高(BLEU 值从 18.77 提高到了 24.93)。这表明了 BPE 机制和共享词表的操作是可以有效地解决数据稀疏性的问题。表中也给出了模型的参数数量对比,可以发现加入这两者操作之后,不但没有增加模

型的参数,而且很大程度上降低了参数的数量,提高了模型的训练效率。

3.4 实验结果

本文对比了三种类型的模型:(1)基于 seq2seq 的模型;(2)基于 graph2seq 的模型;(3)融合外部语料训练的模型。

表 1 给出了模型在两个语料集和融入外部语料的所有实验结果。本文的单个模型性能在 LDC2015E86 和 LDC2017T10 语料集上分别达到了 25.50 和 27.43 的 BLEU 值。在相同的语料集下,本文比传统基于 Bi-LSTM 的 seq2seq 方法(Konstas 等人^[4])高出了 3.5 个 BLEU 值。与图模型相比,本文模型的性能也具有明显的提升。当前最高的模型性能是 Guo 等^[8]人提出的深度连接图卷积网络,但在基于句子级的评测上,本文模型的得分显著的超越他们,在 CHRF++分数上本文取得了 4.55 个点的领先。这说明了本文模型生成的结果相比于他们更贴近于人工的评测,更加贴近真实语义。

遵循着 Konstas 等^[4]人的工作，本文也评估了模型在融入外部语料情况下的性能。当且仅使用 200K 的外部语料时，本文的模型 BLEU 值达到了 37.45，达到了新的最高性能。在使用相同数量级的外部语料情况下，本文分别比 Konstas 等^[4]人，Song 等^[6]人和 Guo 等^[8]人的模型高了 10.05，9.25 和 5.85 个 BLEU 值。特别值得注意的是，本文使用 200K 外部语料得到的结果，相比 Konstas 等^[4]人使用 20M 外部语料依然高出了 3.65 个点，外部语料数量仅仅是他们的一百分之一。这有力的表明了本文模型的优越性。

3.5 AMR 图大小的影响

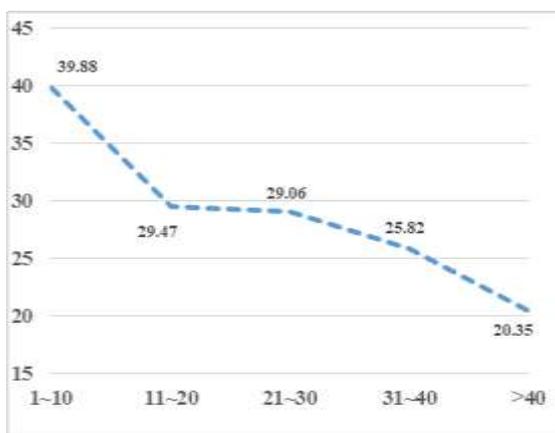


图 3 输入不同 AMR 图大小的模型性能 (BLEU 值)

本文的模型依然是基于 seq2seq 框架，所以为了适应此框架，需要将 AMR 图线性化之后用作输入。这就可能导致在线性化的过程中损失图结构信息。为了证实这个猜想，本文对 AMR 图根据含有不同节点的数量进行了分组，并且在 LDC2017T10 的测试集上各自评估了它们的性能。图 3 以折线图的形式将实验结果展现出来，从图中可以看出：当节点数量在 11 到 30 之间的时候，性能是十分稳定的。但是随着 AMR 图大小的增大，模型的性能持续下降，这种现象也经常出现在神经机器翻译的任务当中。经过分析本文认为对于大规模的 AMR 图性能差主要是由以下原因造成：首先，seq2seq 模型在解析较长的源端句子时，往往较早的预测

出目标端句子的终止符，过早的结束生成。这个问题在神经机器翻译的研究中曾被广泛的讨论过。其次，当图的大小越大，线性化之后损失的结构关系越多，特别是对于图中存在大量的重入节点（有两个及两个以上的父亲节点）情况下尤为严重。这些都增加了模型的学习难度，同时也为我们的下一步工作提供了方向。

3.6 实例分析

为了更好的说明本文模型的优越性，本文给出了一些实例分析，并且选取的是和 Song 等^[6]人以及 Damonte 等^[7]人论文中提出的同样样例进行分析对比。表 3 中实例 (1) 是 Song 等人论文中的实例，其余的是 Damonte 等^[7]人论文中提出的实例。两者都是基于 graph2seq 的模型。需要注意的是，以上模型均未使用外部无标签外部语料。

在实例 (1) 中，我们可以看到传统的 seq2seq 模型并不能通过“p/provide-01”产生名词“agreement”。图模型也错误的判断了 staff, research center 和 funding 它们三者之间的关系。因为图模型只能直接获取到与其相邻节点之间的关系，当图结构复杂分支多的时候，就会导致结构信息的大量丢失。而本文的模型可以精确的识别出它们三者之间的关系。在实例 (2) 中，AMR 图的结构简单，图模型和本文的模型翻译的效果都可以。但是传统的 seq2seq 方法却不能正确的翻译出“he finds out”。图模型和本文的模型都没有正确的翻译正确时态，这是由于当一个句子被解析成 AMR 图时已经丢失了时态导致的。在实例 (3) 中，只有本文的模型正确翻译出了“people”。实例 (4) 中，所有模型都正确的翻译出了“recommend”这个控制结构。但是本文的模型相比于传统的 seq2seq 来说，没有出现过度翻译的问题。实例 (5) 中图模型和我们都正确的翻译出了短语“tell your ex”，而传统的 seq2seq 并没有。

综上所述，可以注意到相比于传统的 seq2seq，本文的模型具有绝对的优势。与图模型相比，本文的模型能够获取到更多的图

表 3 实例分析

<p>(1) (p/ provide-01</p> <p>:ARG0 (a/ agree-01)</p> <p>:ARG1 (a2/ and</p> <p>:op1 (s/ staff</p> <p>:prep-for (c/ center</p> <p>:mod (r/ research-01)))</p> <p>:op2 (f/ fund-01</p> <p>:prep-for c)))</p> <p>Ref: the agreement will provide staff and funding for the research center .</p> <p>Seq: agreed to provide research and institutes in the center .</p> <p>G2S: the agreement provides the staff of the research center and the funding .</p> <p>本文: the agreement provides staff for research centers and funding for center .</p>
<p>(2) Ref: i dont tell him but he finds out ,</p> <p>Seq: i did n't tell him but he was out .</p> <p>G2S: i do n't tell him but he found out .</p> <p>本文: i do n't tell him but he found out .</p>
<p>(3) Ref: if you tell people they can help you ,</p> <p>Seq: if you tell him , you can help you !</p> <p>G2S: if you tell them , you can help you .</p> <p>本文: if you tell people , you might help you !</p>
<p>(4) Ref: i 'd recommend you go and see your doctor too .</p> <p>Seq: i recommend you go to see your doctor who is going to see your doctor .</p> <p>G2S: i recommend you going to see your doctor too .</p> <p>本文: i would recommend going to see your doctor too .</p>
<p>(5) Ref: tell your ex that all communication needs to go through the lawyer .</p> <p>Seq: tell that all the communication go through lawyer .</p> <p>G2S: tell your ex the need to go through a lawyer .</p> <p>本文: tell your ex you need to go all the communication by lawyer .</p>

结构信息，特别是对于那些规模复杂，分支较多的 AMR 图。

4 结论与未来工作

本文第一次将 Transformer 模型引入了 AMR-to-text 生成的任务当中。主要的动机是为了更好的获取到 AMR 图结构信息，克服传统 seq2seq 模型 (Bi-LSTM) 中存在的丢失大量非局部信息的问题，以及图模型在不相邻节点之间信息损失的问题。在两份标准语料集上都取得了当前最高的结果。特别在外部语料的实验中，本文的模型使用最少

量级的语料，达到了目前最高的性能。

在未来的工作中，我们准备以此模型当作基准模型，尝试去解决 AMR 图线性化时丢失图结构信息的问题。

参考文献

- [1] Pourdamghani N, Knight K, Hermjakob U. Generating english from abstract meaning representations[C]//Proceedings of the 9th international natural language generation conference. 2016: 21-25.
- [2] Song L, Peng X, Zhang Y, et al. AMR-to-text

- generation with synchronous node replacement grammar[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017: 7–13.
- [3] Flanigan J, Dyer C, Smith N A, et al. Generation from abstract meaning representation using tree transducers[C]//Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Human language technologies. 2016: 731-739.
- [4] Konstas I, Iyer S, Yatskar M, et al. Neural amr: Sequence-to-sequence models for parsing and generation[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017: 146–157.
- [5] Beck, D., Haffari, G., Cohn, T. Graph-to-Sequence Learning using Gated Graph Neural Networks[C]//Proceedings of 56th Annual Meeting of the Association for Computational Linguistics. 2018: 273–283.
- [6] Song L, Zhang Y, Wang Z, et al. A graph-to-sequence model for AMR-to-text generation[C]//Proceedings of 56th Annual Meeting of the Association for Computational Linguistics. 2018: 1616–1626.
- [7] Damonte M, Cohen S B. Structural Neural Encoders for AMR-to-text Generation [C]//Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2019: 3649–3658.
- [8] Guo Z, Zhang Y, Teng Z, et al. Densely Connected Graph Convolutional Networks for Graph-to-Sequence Learning[C]//Proceedings of Transactions of the Association for Computational Linguistics. 2019: 297–312.
- [9] Banarescu L, Bonial C, Cai S, et al. Abstract meaning representation for sembanking[C] //Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse. 2013: 178-186.
- [10] Tamchyna A, Quirk C, Galley M. A discriminative model for semantics-to-string translation[C] //Proceedings of the 1st Workshop on Semantics-Driven Statistical Machine Translation (S2MT 2015). 2015: 30-36.
- [11] Mitra, A., Baral, C. Addressing a question answering challenge by combining statistical methods with inductive rule learning and reasoning[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016: 2779–2785.
- [12] Li, X., Nguyen, T. H., et al. Improving Event Detection with Abstract Meaning Representation[C]//Proceedings of the First Workshop on Computing News Storylines. 2015: 11–15.
- [13] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[C]//Proceedings of Neural Information Processing Systems, 2017: 5998–6008.
- [14] Sennrich R, Haddow B, Birch A. Neural machine translation of rare words with subword units[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2015: 1715–1725.
- [15] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Proceedings of Neural Information Processing Systems. 2014: 3104-3112.
- [16] Cao K, Clark S. Factorising AMR generation through syntax[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019: 2157-2163.
- [17] Sharif Razavian A, Azizpour H, Sullivan J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2014: 806-813.
- [18] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [19] Zhang, Y., Liu, Q., Song, L. Sentence-State LSTM for Text Representation[C]//Proceedings of

- 56th Annual Meeting of the Association for Computational Linguistics. 2018: 317-327.
- [20] Barzdins G, Gosko D. Riga at semeval-2016 task 8: Impact of smatch extensions and character-level neural translation on amr parsing accuracy[C]//Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 2016: 1143-1147.
- [21] van Noord R, Bos J. Neural semantic parsing by character-based translation: Experiments with abstract meaning representations[J]. Computational Linguistics in the Netherlands Journal, 2017: 93-108.
- [22] Gage P. A new algorithm for data compression[J]. The C Users Journal, 1994, 12(2): 23-38.
- [23] Flanigan J, Dyer C, Smith N A, et al. Generation from abstract meaning representation using tree transducers[C]//Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Human language technologies. 2016: 731-739.
- [24] Ge D, Li J, Zhu M, et al. Modeling Source Syntax and Semantics for Neural AMR Parsing[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019: 4975 - 4981.
- [25] Kingma D P, Ba J. Adam: A method for stochastic optimization[C]// Proceedings of the International Conference on Learning Representations. 2015.
- [26] 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.
- [27] Banerjee S, Lavie A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments[C]//Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. 2005: 65-72.
- [28] Popović M. chrF++: words helping character n-grams[C]//Proceedings of the second conference on machine translation. 2017: 612-618.

作者联系方式

姓名: 朱杰

地址: 江苏省苏州市姑苏区苏州大学

邮编: 215006

电话: 19951561121

电子邮箱: zhujie951121@gmail.com

姓名: 李军辉

地址: 江苏省苏州市姑苏区苏州大学

邮编: 215006

电话: 13812696576

电子邮箱: lijunhui26@gmail.com

姓名: 钱龙华

地址: 江苏省苏州市姑苏区苏州大学

邮编: 215006

电话: 13306208165

电子邮箱: qianlonghua@suda.edu.cn