



融入句法感知表示进行句法增强的语义解析

谢德峰, 吉建民*

(中国科学技术大学 计算机科学与技术学院, 合肥 230027)

(* 通信作者电子邮箱 jianmin@ustc.edu.cn)

摘要:在自然语言处理(NLP)中,句法信息是完整句子中词汇与词汇之间的句法结构关系或者依存关系,是一种重要且有效的参考信息。语义解析任务是将自然语言语句直接转化成语义完整的、计算机可执行的语言。在以往的语义解析研究中,少有采用输入源的句法信息来提高端到端语义解析效率的工作。为了进一步提高端到端语义解析模型的准确率和效率,提出一种利用输入端句法依存关系信息来提高模型效率的语义解析方法。该方法的基本思路是先对一个端到端的依存关系解析器进行预训练;然后将该解析器的中间表示作为句法感知表示,与原有的字词嵌入表示拼接到一起以产生新的输入嵌入表示,并将得到的输入嵌入表示用于端到端语义解析模型;最后采用转导融合学习方式对模型进行融合。实验对比了所提模型和基准模型Transformer以及过去十年的相关工作。实验结果表明,在ATIS、GEO、JOBS数据集上,融入依存句法信息感知表示以及转导融合学习的语义解析模型分别实现了89.1%、90.7%、91.4%的最佳准确率,全面超过了Transformer,验证了引入句法依存关系信息的有效性。

关键词:句法感知;语义解析;深度学习;自然语言处理;语言模型

中图分类号:TP391 **文献标志码:**A

Syntax-enhanced semantic parsing with syntax-aware representation

XIE Defeng, JI Jianmin*

(School of Computer Science and Technology, University of Science and Technology of China, Hefei Anhui 230027, China)

Abstract: Syntactic information, which is syntactic structure relations or dependency relations between words of a complete sentence, is an important and effective reference in Natural Language Processing (NLP). The task of semantic parsing is to directly transform natural language sentences into semantically complete and computer-executable languages. In previous semantic parsing studies, there are few efforts on improving the efficiency of end-to-end semantic parsing by using syntactic information from input sources. To further improve the accuracy and efficiency of the end-to-end semantic parsing model, a semantic parsing method was proposed to utilize the source-side dependency relation information of syntax to improve the model efficiency. As the basic idea of the method, an end-to-end dependency relation parser was pre-trained firstly. Then, the middle representation of the parser was used as syntax-aware representation, which was spliced with the original word embedding representation to generate a new input embedding representation, and this obtained input embedding representation was used in the end-to-end semantic parsing model. Finally, the model fusion was carried out by the transductive fusion learning. In the experiments, the proposed model was compared with the baseline model Transformer and the related works in the past decade. Experimental results show that, on ATIS, GEO and JOBS datasets, the semantic parsing model integrating dependency syntax-aware representation and transductive fusion learning achieves the best accuracy of 89.1%, 90.7%, and 91.4% respectively, which exceeds the performance of the Transformer. It verifies the effectiveness of introducing the dependency relation information of syntax.

Key words: syntax-aware; semantic parsing; deep learning; Natural Language Processing (NLP); language model

0 引言

伴随着21世纪第二个十年期间人工智能领域的发展与各种深度学习模型的提出,人工智能已经成为当今计算机学科中非常重要的一个子学科,成为了21世纪最火热的话题。与此同时,机器人作为人工智能领域的一个子学科,综合了计算机学科中多个研究方向,也成为了许多学术团队和科技公司正在大力研究的对象。随着经济的发展,生活辅助型智能

应用越来越多,给人们带来了非常大的便利,智能机器人也逐步进入市场,为人们提供简单的劳动服务。人们希望能用自然语言直接与机器人或者应用进行沟通,从而让机器人或者应用更好地服务人们,为人们带来便利。

文献[1]中指出,自然语言理解是机器人或者应用能够提供智能服务的关键技术之一。语义解析(semantic parsing)的定义就是这样:将一个自然语言语句转化成机器能够理解或执行的语言,比如逻辑表达形式或可执行的程序,从

收稿日期:2020-11-27;修回日期:2021-01-15;录用日期:2021-01-19。

基金项目:科技创新2030—“新一代人工智能”重大项目(2018AA000500);广东省科技计划项目(2017B010110011)。

作者简介:谢德峰(1994—),男,广东汕头人,硕士研究生,主要研究方向:语义分析、机器阅读理解、命名实体识别;吉建民(1984—),男,甘肃陇西人,副教授,博士,CCF会员,主要研究方向:自然语言处理、深度强化学习、认知机器人。



而使计算机理解自然语言。语义解析作为自然语言理解的核心,在很多任务中都是关键步骤。语义解析是让机器人理解自然语言的关键步骤。

随着深度学习的发展,许多团队也试着利用深度学习模型来完成语义解析的任务。但由于现有语义解析数据集的规模较小,远远不能满足深度学习所需要的数据量,且构建和标注新的数据集非常困难,因为标注形式不是简单的类别而是一串逻辑符号组成的表达式,标注人员需要专业知识背景,因此,数据集少且规模小成为了制约语义解析发展的一个问题。此外,语义解析还对模型输出的精度要求高,输出序列只要有其中一个字符解析错误,那么一整串序列输出都是不正确且不可用的。同时,语义解析过程中还存在两大难题:其中一个问题是长期依赖问题,即需要生成的某个元素依赖于另一个遥远的元素;另一个问题是结构建模问题,即所要生成的序列需要符合逻辑表示的语法要求,该语法包含了丰富的结构信息。如何在上述一系列制约因素以及精度要求之下进一步提高语义解析模型准确率及效率,成为亟须研究和解决的难题。

句法解析(syntactic parsing)是自然语言处理中一个重要的任务,其目标是分析句子的语法结构并将其表示为容易理解的结构(通常是树形结构)。句法解析与语义解析是两个完全不同的任务,二者的实现方式也截然不同:句法解析的目标是解析出句子中词汇与词汇之间的句法关系(包括结构关系和依存关系);语义解析的目标在于解析句子的内在含义并将该含义抽象成逻辑表示。一般而言,句法解析模型生成图形式的句法关系,而语义解析生成序列形式的逻辑语句。句法解析得到的句法关系可作为特征信息用于辅助语义解析任务。句法解析是所有工具性自然语言处理(Natural Language Processing, NLP)任务中较为高级、复杂的一种任务。文献[2]阐述了句法解析可进一步分为句法结构解析(syntactic structure parsing)和依存关系解析(dependency parsing)两种,其中:句法结构解析是通过分析语句中词与词之间的相互联系,相互作用的方式,得到的整个涵盖整个句子全局的句法结构树;而依存关系解析则是根据语法依存理论,分析语言单位内成分之间的依存关系从而揭示其句法依存结构。

以往的基于深度学习的语义解析方法大多是以序列到序列(Seq2Seq)模型^[3-4]为基础。在这些研究中,大多没有直接用到输入源的自然语言句子的句法特性,而是将识别句法特性的任务统一地交给了端到端神经网络去自主处理,但神经网络很难能在没有相关标注信息的情况下自主识别出输入句子的句法特性。利用自然语言本身的句法特征信息可以为模型增加可参考的输入信息,有助于缓解语义解析中语言结构建模和数据集规模小的瓶颈问题。为了进一步提高语义解析模型的准确率和效率,本文采用一种利用句法信息来增强模型效率的语义解析方法,即利用句法解析任务来辅助语义解析任务。该方法利用输入句子的句法依存关系来增强端到端语义解析模型的效率。基本思路是在已有端到端语义解析模型的基础上,先对一个端到端的依存关系解析器进行预训练,将该依存关系解析器的中间表示作为句法感知表示,与原有的字词嵌入(word embedding)表示拼接到一起,产生新的输入嵌入表示,而后再将该新的输入嵌入表示用于端到端语义解析模型,以此来提高语义解析模型的效率。

近几年也有一些工作开始尝试利用句法信息来增强端到

端语义解析,例如文献[5-6]采用树形长短期记忆(Tree Long Short-Term Memory, Tree-LSTM)单元来处理句法结构信息,但本文提出的融入句法感知表示进行句法增强的语义解析更加简单且有效。文献[7]采用了融入句法感知表示的思路,受其启发,本文工作将该思路应用于语义解析领域。本文实验使用基于Transformer的Seq2Seq模型作为基准模型,采用转导融合学习方式进行模型融合。实验同时对比了最近10年的语义解析研究工作。最终实验结果表明,在ATIS、GEO、JOBS数据集上,本文采用的句法增强语义解析方法不仅全面超过了基准水平,还在大部分指标上超越了其他相关工作。

1 相关工作

最传统的语义解析包括模板匹配、手写语法规则进行推导等方法。虽然有些工作已经能达到非常高的准确率,但是都需要进行大量的前期工程,依赖于手工设计的规则,通用性较差。

统计语义解析方法主要是运用了机器学习方法,依赖于有标注的自然语言和逻辑表达式配对的数据集。文献[8-10]中依赖了自然语言语句的依存关系解析结果,用lambda算子或者组合范畴语法分析器来将语句的依存结构转化为逻辑形式;文献[11]中没有用到手工规则或者组合范畴语法,而是直接采用了较为暴力的方法,通过连接(Join),求交(Intersection)和聚合(Aggregate)三种操作,以及提出的桥接操作自底向上地对语法树的节点进行两两合并,从而在根节点得到逻辑形式;文献[12]将物理特征分类器和基于概率范畴语法的语言模型两者进行联合学习,以此让语言模型能够构造组合意义表示;文献[13]中采用了组合类别语法归纳来处理语义解析任务;文献[14]通过构建一个词法语法分解词典来建模词类之间的词法、语法和语义,以此提升组合范畴语法的效率。

随着深度学习的发展,许多基于神经网络Seq2Seq模型或其衍生模型大量涌现。深度学习的优点在于其在不需要特征工程的情况下,取得了不亚于手工转换方法的准确率,但同时也带来了难以解释模型的缺点。文献[15]中采用基于LSTM单元^[16-17]和基于注意力机制^[4,18]的Seq2Seq模型来解决语义解析问题;同时文献[15]还提出了统一基于LSTM和注意力机制的序列到树(Seq2Tree)模型来解决语义解析中的结构建模问题,提高语义解析效率。自此,许多研究团队都以文献[15]为基础,在编码器端或者解码器端进行模型上的改进。文献[19]在编码器端融入了自然语言的句法信息,考虑了单词顺序、依存结构特征和句法成分特征,构建了一个图-序列(Graph2Seq)模型。文献[20]则在解码器端生成了一个语义图,构建了一个序列-语义图模型。文献[21]虽然是以Seq2Seq模型为基础,但是根据结构化查询语言(Structured Query Language, SQL)的特点,在解码器端分成三个频道并加入了一个门单元。三个频道分别预测数据表中列名称、表中单元格名称以及SQL语法关键字;而门单元则是一个控制开关,在每个时间节点预测应该选择哪个频道的预测结果作为输出。文献[22]则用了两层的编码器+解码器模型(encoder-decoder)来实现从粗略到精细的语义解析过程:第一层先生成草稿,第二层再生成其中的变量作为补充。TRANX^[23]借鉴了抽象语法树(Abstract Syntax Tree, AST)的思想,根据目标



语言的语法构建规约文法,基于该文法可以将生成目标表示为语法树,然后实现自顶向下的语法树生成系统。文献[24]提出了一种自适应的解码方式来帮助解码器自主学习要生成的逻辑形式结构。

近年来随着预训练语言模型的兴起,很多领域都开始注意到了众多预训练语言模型的基础模型 Transformer^[25]。Transformer 具有媲美 LSTM 的处理远距离依赖的能力,同时也具备与卷积神经网络相似的并行可行性,它在处理大规模数据集时比 LSTM 要更具鲁棒性,不容易过拟合。TreeGen^[26]借助 Transformer^[25]作为解码器(decoder),同时也采用基于 Transformer 改造的可用于语义解析的 AST 阅读器来编码抽象句法树。

2 方法与模型

2.1 基于深度学习的依存关系解析器

为了在语义解析模型中融入依存句法信息感知表示,首先需要预训练一个依存关系解析器。本文所使用的句法信息为依存关系信息,依存关系解析是根据语法依存理论,解析语言单位内成分之间的依存关系从而揭示其句法依存结构,依存关系树则描述了原句中词与词之间的依存关系。如图 1 所示为一个句子中的依存关系信息示例。

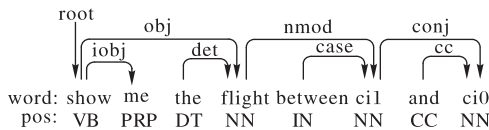


图 1 依存关系信息示例

Fig. 1 Example of dependency relation information

依存关系解析器的作用是根据输入句子及其词性标注标记出其依存关系,从而获得该句子的依存关系解析结构图。依存关系解析器本质上是一个序列到图的生成器,该依存关系解析过程可如图 2 所示。其中, $r_1 \sim r_8$ 为编码器生成的中间表示。

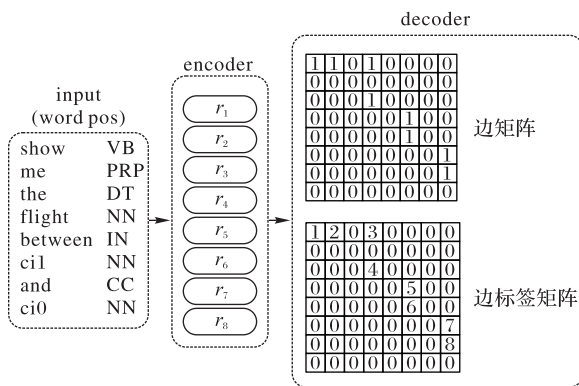


图 2 依存关系解析的解析过程

Fig. 2 Process of dependency relation parsing

本文所采用的依存关系解析器源自于文献[27],该解析器已被证明了相较于文献[28]中所提出的模型更加简单却更高效,因此采用该依存关系解析器提取输入句子中的依存关系信息。如图 3 所示,该依存关系解析器由双向 LSTM、四个全连接层和两个特殊的计算参数矩阵所构成。可以将该模型看作是编码器(encoder)和 decoder 两个部分,嵌入层和 LSTM 层为 encoder 部分,其余上层为 decoder 部分。首先,模型的输

入是单词和单词各自的词性标注信息(part-of-speech),将输入的句子词嵌入(word embedding)和词性标注信息嵌入(Part-Of-Speech embedding, POS embedding)拼接起来作为输入,然后用多层双向长短期记忆层(Bi-directional LSTM, BiLSTM)处理该输入,得到中间表示 R 。如式(1)~(2):

$$x_i = e_i^{(\text{word})} \oplus e_i^{(\text{tag})} \quad (1)$$

$$R = \text{BiLSTM}(X) \quad (2)$$

其中: $e^{(\text{word-dependency})}$ 、 $e^{(\text{pos})}$ 为依存关系解析器端的词嵌入和标注信息嵌入。

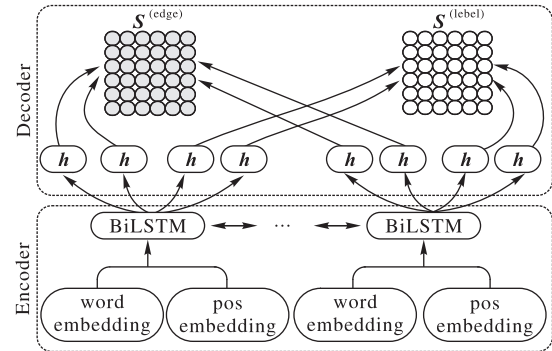


图 3 依存关系解析器模型结构

Fig. 3 Structure of dependency relation parser

随后,用 4 个不同的全连接层分别处理中间表示 R ,得到 4 个表示 $h_i^{(\text{edge-head})}$ 、 $h_i^{(\text{edge-dep})}$ 、 $h_i^{(\text{label-head})}$ 、 $h_i^{(\text{label-dep})}$,如式(3)~(6)所示,其含义与文献[27]中含义相同。这一步的操作的目的是在保证 LSTM 解析能力不损失的情况下对 R 进行降维处理,达到防止过拟合的目的。

$$h_i^{(\text{edge-head})} = \text{FNN}^{(\text{edge-head})}(r_i) \quad (3)$$

$$h_i^{(\text{label-head})} = \text{FNN}^{(\text{label-head})}(r_i) \quad (4)$$

$$h_i^{(\text{edge-dep})} = \text{FNN}^{(\text{edge-dep})}(r_i) \quad (5)$$

$$h_i^{(\text{label-dep})} = \text{FNN}^{(\text{label-dep})}(r_i) \quad (6)$$

接着,将上一步得到边以及边上标签的表示进行双衍射处理^[28],如式(7)~(8)所示。双衍射公式计算过程如式(9)~(10)所示。

$$s_{i,j}^{(\text{edge})} = \text{Biaff}^{(\text{edge})}(h_i^{(\text{edge-dep})}, h_j^{(\text{edge-head})}) \quad (7)$$

$$s_{i,j}^{(\text{label})} = \text{Biaff}^{(\text{label})}(h_i^{(\text{label-dep})}, h_j^{(\text{label-head})}) \quad (8)$$

$$\text{Bilin}(x_1, x_2) = x_1^T U x_2 \quad (9)$$

$$\text{Biaff}(x_1, x_2) = x_1^T U x_2 + W(x_1 \oplus x_2) + b \quad (10)$$

值得注意的是:当双衍射处理是为了处理是否有边存在的,即结果为得到 $S^{(\text{edge})}$ 时, U 的维度为 $(d \times 1 \times d)$, d 表示 $h_i^{(\text{edge-head})}$ 、 $h_i^{(\text{edge-dep})}$ 、 $h_i^{(\text{label-head})}$ 、 $h_i^{(\text{label-dep})}$ 的维度;当双衍射处理是为了处理是否有边上标签信息,即结果为得到 $S^{(\text{label})}$ 时, U 的维度为 $(d \times c \times d)$,其中 c 是所有边可能标签的数量总和。最后得到的边矩阵 $S^{(\text{edge})}$ 中, $s_{i,j}^{(\text{edge})}$ 表示节点 i 指向节点 j 的边存在的概率,而边标签矩阵 $S^{(\text{label})}$ 中, $s_{i,j}^{(\text{label})}$ 表示节点 i 指向节点 j 的边(若存在)的所有可能标签的概率。该过程如式(11)~(12)所示:

$$y_{i,j}^{(\text{edge})} = \{\text{sigmoid}(s_{i,j}^{(\text{edge})}) \geq 0\} \quad (11)$$

$$y_{i,j}^{(\text{label})} = \arg \max(\text{soft max}(s_{i,j}^{(\text{label})})) \quad (12)$$

与文献[27]相同,本文通过计算预测分布与真实标签分布的损失,然后通过反向传播梯度下降来训练模型。用



softmax cross-entropy 来计算边标签预测与边真实标签之间的损失,用 sigmoid cross-entropy 来衡量边存在概率与边存在真实标签之间的损失。然后引入超参数 λ 来平衡这两个损失函数的权重。过程如式(13)~(15)所示:

$$Loss^{edge} = \text{cross-entropy}(y_{i,j}^{(edge)}, y_{i,j}^{(edge-groundtruth)}) \quad (13)$$

$$Loss^{label} = \text{cross-entropy}(y_{i,j}^{(label)}, y_{i,j}^{(label-groundtruth)}) \quad (14)$$

$$Loss = \lambda Loss^{label} + (1 - \lambda) Loss^{edge} \quad (15)$$

2.2 基于 Transformer 的语义解析模型

在语义解析领域,以往的多数研究都是以 LSTM 和注意力机制为基础 Seq2Seq 模型。近年来随着预训练语言模型的兴起,Transformer^[25] 凭借其处理远距离关系依赖的能力和与卷积神经网络相似的并行能力而受到广泛关注。本文也采用 Transformer 作为基础的语义解析模型。

Transformer 引入了自注意力机制和多头自注意力结构。和一般的 Seq2Seq 模型一样,本质上是一个 encoder-decoder 的结构。encoder 和 decoder 都由多个单独的处理单元堆叠而成,单元与单元之间采用层归一化和残差链接。如式(16)所示:

$$\begin{aligned} \text{sub_layer_output} = & \\ & \text{LayerNorm}(\text{sub_layer_input} + \\ & (\text{SubLayer}(\text{sub_layer_input}))) \end{aligned} \quad (16)$$

Transformer 由多头自注意力层(multi-head self-attention)和位置前馈网络层(position-wise feed-forward networks)组成,其中多头自注意力模块有多个缩放点乘运算构(scaled dot-product)。上述多头自注意力层和位置前馈网络层的计算过程如式(17)~(19)所示:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) W^o \quad (17)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (18)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (19)$$

decoder 部分跟 encoder 相比,多了个一层用于处理 encoder 所输出中间表示的多头注意力层,并且在自身的自注意力层加入了掩码机制。Transformer 还采用三角函数还表示位置编码(position encode),以表明输入节点的位置信息。

2.3 融入句法信息感知的语义解析模型

与文献[7]中的思路相同,为了融入句法信息感知,需要先预训练 2.1 节所示的依存关系解析器模型,让依存关系解析器有了初步理解句子依存关系的能力。而后将预训练好的依存关系解析器模型的 encoder 部分的输出,也就是式(2)中所示的 R ,作为所要用来提供句法信息的句法感知表示(syntax-aware representation)。

如式(20)所示,该过程是将得到的句法感知表示 R 与语义解析模型 encoder 端的词嵌入拼接到一起,加上位置编码(position encode),再将其作为 Transformer-encoder 的输入,其中 $e^{(\text{transformer-word-embed})}$ 是 encoder 端的词嵌入, $e^{(\text{position-encode})}$ 是位置编码。

$$\text{Memory} = \text{Transformer-Encoder}((R \oplus e^{(\text{transformer-word-embed})}) + e^{(\text{position-encode})}) \quad (20)$$

decoder 部分保持与 Transformer 的 decoder 部分一致。如图 4 所示,左侧模型为用来提取依存关系信息的依存关系解析器,右侧模型为 Transformer 模型。可以看出,原本的 Transformer 只需要输入词嵌入+位置编码,采用本文方法改进后,Transformer 的输入是词嵌入与依存关系解析器的中间表示 R 拼接后的向量+位置编码,这样的改进旨在让带有依存句法信息的中间表示 R 以嵌入(embedding)的形式,让模型可以通过解析该中间表示 R 得到句法信息。

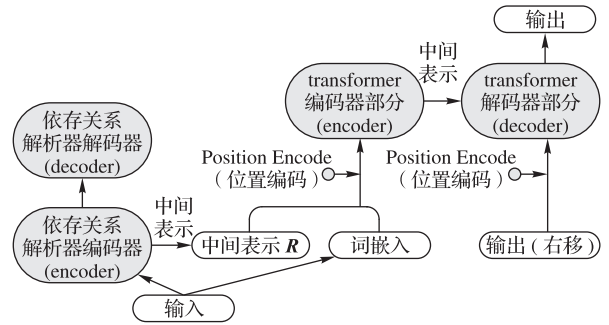


图4 融入句法信息感知后的 Transformer 整体结构

Fig. 4 Overall structure of Transformer after fusing syntax-aware representation

2.4 转导融合学习

模型融合是提高机器学习任务准确性的一种常见实践。由于语义解析任务的数据集规模小,标注成本昂贵,常规的集成方法仅能对语义解析任务带来微不足道的改进。为了探索更好的模型融合效果,本文将转导融合学习用于语义解析任务。转导融合学习是由文献[29]提出的一种简单有效的模型融合方式,它使用若干个单独的模型将源测试集的输入数据转换到目标语言空间,然后用单模型在转换后的合成数据上进行微调。相较于传统模型融合方式,不需要多个子模型同时在线预测,对于像 Seq2Seq 这种推理过程为贪婪搜索或束搜索(beam search)的模型,转导融合学习算法过程简单、高效,但缺点是在融合过程中需要用到待预测的不包含标注的输入信息。转导融合学习可以看作是一种半监督学习方式。

转导融合学习使用若干子模型来对验证集或测试集的输入进行预测。式(21)~(23)表示验证集,测试集和已在训练集上训练得到的 K 个子模型。采用子模型对验证集和测试集的输入部分进行推理,得到基于子模型的合成数据,如式(24)~(25)所示。选择式(23)中的任意一个子模型,用合成数据集对其进行微调,在微调过程中,用验证集进行验证或采用交叉验证选出合适参数的模型,最终得到融合好的模型。

$$D_{\text{valid}} = \{\bar{x}_i, \bar{y}_i\}_{i=1}^{N_{\text{valid}}} \quad (21)$$

$$D_{\text{test}} = \{x_i^*, y_i^*\}_{i=1}^{N_{\text{test}}} \quad (22)$$

$$\text{sub_models} = f_1, f_2, \dots, f_K \quad (23)$$

$$D_{\text{v-synthetic}} = \{(\bar{x}_i, f_k(\bar{x}_i)) | \bar{x}_i \in D_{\text{valid}}, k \in \{1, 2, \dots, K\}\} \quad (24)$$

$$D_{\text{t-synthetic}} = \{(x_i^*, f_k(x_i^*)) | x_i^* \in D_{\text{test}}, k \in \{1, 2, \dots, K\}\} \quad (25)$$

3 实验和结果分析

3.1 数据集

实验使用 ATIS、GEO、JOBS 三个语义解析数据集,并采用



了文献[15]提供的预处理版本,这些数据集的输入是自然语言描述,而输出则是lambda演算形式的代码序列。在实验中遵循了这些数据集的标准train-dev-test划分,数据集的统计信息在表1中列出。

表 1 数据集统计信息
Tab. 1 Statistics of datasets

数据集	样本数			平均输入长度	最大输入长度	平均标注长度	最大标注长度
	训练集	验证集	测试集				
ATIS	4 434	491	448	10.5	44	28.2	94
GEO	600	—	280	7.4	23	19.1	51
JOBS	500	—	140	8.7	22	10.0	51

3.2 实验设置和细节

3.2.1 评价指标

对 ATIS 和 GEO 数据集采用完全匹配(Exact Match)和树型匹配(Tree Match)这两个基于准确率的指标来度量最终模型结果的好坏。Exact Match是只有当模型预测序列的单词及位置均与测试给定的标注完全一致时,才认为该例的预测结果是正确的。而Tree Match则遵循文献[16]中的度量标准,文献[16]认为预测结果只需与标注结果含义相同即可认为是正确的预测结果,采用Tree Match意味着当预测结果序列的树形结构表示与标注结果目标逻辑形式的树形结构相同时即可认为该例的预测结果是正确的。对于JOBS数据集,本实验只采用Exact Match进行最终结果度量。

3.2.2 实验细节参数设置

实验采用Stanford corenlp-4.1.0^[30]对数据集的输入部分进行分词(tokenization)、词性标注(part-of-speech tagging)和依存关系解析,将各个数据集的训练集部分输入部分的分词结果和词性标注结果作为2.1节中的依存关系解析器的输入,将依存关系解析结果作为2.1节中的依存关系解析器的标注数据,以此来预训练该依存关系解析器。

对于两个模型,实验均采用AdamW作为优化器,AdamW的参数betas为(0.9, 0.999),eps为1E-8,weight decay coefficient为1E-2,实验采用提前停止(early stop)来防止模型过拟合。对于GEO、JOBS,实验利用交叉验证来寻找合适的模型参数,而对于ATIS则是在其验证集上验证。实验所使用的依存关系解析器与基于Transformer语义解析模型的各项参数设置如表2~3所示,表中DR为丢弃率(Dropout Rate),DI为维度(Dimension)。

表 2 依存关系解析器的参数设置
Tab. 2 Parameter setting of dependency relation parser

参数	值	参数	值
word embedding DI	128	pos embedding DI	128
BiLSTM hidden DI	256	BiLSTM output DI	512
$h^{(\text{edge-head})}$ DI	256	$h^{(\text{edge-dep})}$ DI	256
$h^{(\text{label-head})}$ DI	256	$h^{(\text{label-dep})}$ DI	256
word embedding DR	0.2	pos embedding DR	0.2
BiLSTM output DR	0.25	$h^{(\text{label-dep})}$ DR	0.45
$h^{(\text{edge-head})}$ DR	0.45	$h^{(\text{edge-dep})}$ DR	0.45
$h^{(\text{label-head})}$ DR	0.45	λ	0.025
batch size	16	learning rate	1E-4

表 3 语义解析器的参数设置

Tab. 3 Parameter setting of semantic parser

参数	值	参数	值
word embedding DI	512	Feedforward DI	2 048
encoder layer	2	decoder layer	3
head number	8	attention DI	512
activation	ReLU	dropout rate	0.1
batch size	16	learning rate	1E-4

3.3 实验结果

本文在不同数据集上做了四组实验:

实验 1 基于Transformer的语义解析模型作为实验的基准模型(baseline);

实验 2 在实验1的基础上加入了转导融合学习;

实验 3 本文所提出的融入依存关系信息感知表示的语义解析模型;

实验 4 在实验3的基础上加入了转导融合学习。

融入依存关系信息感知表示的语义解析模型为实验主要模型,同时,采用了转导融合学习来进一步提高各自的效率,以此来对比两个模型各自融合前和融合后的效果差异。在ATIS、GEO数据集上的实验结果如表4所示。JOBS数据集上的实验结果如表5所示。外延匹配是一种使用了外部知识库来辅助计算准确率的方式。

3.4 实验结果分析

对比实验1和基于LSTM的Seq2Seq^[15]工作,在三个数据集上,使用基于Transformer的Seq2Seq模型均可以达到比基于LSTM的Seq2Seq模型更好的效果,可以看出Transformer确实对于远距离依赖的解析能力较强。对比实验2和实验1(或对比实验4和实验3)的实验结果,可以看出转导融合学习对于ATIS、GEO、JOBS这类规模较小但标注复杂的语义解析数据集有较好的提升效果。对比实验3和实验1可以看出融入了依存句法信息感知表示的模型比不融入该信息的模型表现更好,说明该本文所提的融入句法信息感知方式是有效的。

对于ATIS数据集,实验2、实验3和实验4在Exact match指标上都超越了之前已有的基于神经网络的工作的最佳结果,而实验4在Tree match指标上达到了与之前最佳结果相同的水准。对于GEO数据集,实验4在Tree match指标上超越了之前已有基于神经网络的工作的最佳结果,而在Exact match指标上表现不好,通过分析可看出,是由于模型的解析能力较强,能较灵活地学到语义解析中的结构建模信息,因而在Tree match指标上表现良好,但在Exact match指标上却表现一般。对于JOBS,由于现有可对比工作较少,只在Exact match指标上进行对比,实验4超越了之前已有工作的最佳结果,而实验3也达到了接近于现有工作最佳水平的结果。

可以看到融入依存句法信息感知表示以及转导融合学习对语义解析模型的有效性。实验结果表明本文提出的融入依存信息感知的句法信息增强模型即使在没有使用转导融合学习的情况下,也能达到与现有工作水平相近甚至在某个指标上可以超越的结果,而加入转导融合学习之后,可以在多个指标数超过其他相关工作。这验证了引入句法依存关系信息的有效性。



表 4 ATIS、GEO 数据集上的准确率
Tab. 4 Accuracy on ATIS and GEO datasets

单位:%
unit:%

方法类别	算法模型	ATIS		GEO	
		Exact match	Tree match	Exact match	Tree match
传统机器学习方法	ZC07 ^[31]	84.6	—	86.1	—
	FUBL ^[32]	82.8	—	88.6	—
	DCS ^[33]	—	—	87.9	—
	KCAZ13 ^[34]	—	—	89.0	—
	WKZ14 ^[14]	91.3	—	90.4	—
	TISP ^[35]	84.2	—	88.9	—
基于神经网络的方法	Seq2Seq ^[15]	—	84.2	—	84.6
	Seq2Tree ^[15]	—	84.6	—	87.1
	JL16 ^[36]	83.3	—	89.3*	—
	TRANX ^[23]	—	86.2	—	88.2
	Coarse2Fine ^[22]	—	87.7	—	88.2
	Seq2Act ^[20]	85.5	—	88.2*	—
	Graph2Seq ^[19]	85.5	—	88.9	—
	AdaNSP ^[24]	—	88.6	—	88.9
	TreeGen ^[26]	—	89.1	—	89.6
	实验 1	85.1	86.2	83.2	87.5
本文方法	实验 2	86.6	87.1	84.6	88.2
	实验 3	86.6	87.7	85.0	87.9
	实验 4	88.4	89.1	86.8	90.7

注:*为外延匹配(Denotation Match)方式计算得到。

表 5 JOBS 数据集上的准确率
Tab. 5 Accuracy on JOBS dataset

单位:%
unit:%

方法类别	算法模型	JOBS
传统机器学习方法	ZC07 ^[31]	79.3
	DCS ^[33]	90.7
	TISP ^[35]	85.0
基于神经网络的方法	Seq2Seq ^[15]	87.1
	Seq2Tree ^[15]	90.0
	Graph2Seq ^[19]	91.2
	实验 1	87.9
	实验 2	88.6
	实验 3	90.7
	实验 4	91.4

4 结语

本文提出了一个融入依存关系信息的端到端句法信息增强语义解析模型,并采用了转导融合学习的方法来融合模型。在 ATIS、GEO、JOBS 数据集上,该句法增强语义解析方法不仅全面超过了基准水平,还在大部分指标上超越了其他相关工作。且即使在没有经过转导融合学习的情况下,也能达到与现有相关工作水平相近甚至在某些指标上超越相关工作的水平。接下来我们将会进一步在语义解析这个方向上探究如何更好地融入句法信息,或寻找更有效的新方法。

参考文献 (References)

- [1] 吉建民. 提高 ASP 效率的若干途径及服务机器人上应用[D]. 合肥:中国科学技术大学, 2010: 2. (JI J M. Several ways to improve efficiency of answer set programming and its application in service robots[D]. Hefei: University of Science and Technology of China, 2010: 2.)
- [2] 柴伟. 短语结构句法分析综述[J]. 电脑知识与技术, 2020, 16(16): 26-27, 30. (CHAI W. A review of the constituency parsing [J]. Computer Knowledge and Technology, 2020, 16(16): 26-27, 30.)

- [3] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. (2016-05-19) [2020-09-10]. <http://www.arxiv.org/pdf/1409.0473.pdf>.
- [4] CHO K, VAN MERRIËNBOER B, GULÇEHRE Ç, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2014: 1724-1734.
- [5] ERIGUCHI A, HASHIMOTO K, TSURUOKA Y. Tree-to-sequence attentional neural machine translation [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2016: 823-833.
- [6] CHEN H D, HUANG S J, CHIANG D, et al. Improved neural machine translation with a syntax-aware encoder and decoder [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2017: 1936-1945.
- [7] ZHANG M S, LI Z H, FU G H, et al. Syntax-enhanced neural machine translation with syntax-aware word representations [C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2019: 1151-1161.
- [8] REDDY S, LAPATA M, STEEDMAN M. Large-scale semantic parsing without question-answer pairs [J]. Transactions of the Association for Computational Linguistics, 2014, 2: 377-392.
- [9] REDDY S, TÄCKSTRÖM O, PETROV S, et al. Universal semantic parsing [C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2017: 89-101.
- [10] REDDY S, TÄCKSTRÖM O, COLLINS M, et al. Transforming dependency structures to logical forms for semantic parsing [J]. Transactions of the Association for Computational Linguistics, 2016, 4: 127-140.
- [11] BERANT J, CHOU A, FROSTIG R, et al. Semantic parsing on freebase from question-answer pairs [C]// Proceedings of the 2013



- Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2013: 1533-1544.
- [12] MATUSZEK C, FITZGERALD N, ZETTLEMOYER L, et al. A joint model of language and perception for grounded attribute learning [C]// Proceedings of the 29th International Conference on Machine Learning. Madison, WI: Omnipress, 2012: 1671-1678.
- [13] ZETTLEMOYER L S, COLLINS M. Learning to map sentences to logical form: structured classification with probabilistic categorical grammars [C]// Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence. Arlington, VA: AUAI Press, 2005: 658-666.
- [14] WANG A, KWIATKOWSKI T, ZETTLEMOYER L. Morpho-syntactic lexical generalization for CCG semantic parsing [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2014: 1284-1295.
- [15] DONG L, LAPATA M. Language to logical form with neural attention [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2016: 33-43.
- [16] MALHOTRA P, VIG L, SHROFF G, et al. Long short term memory networks for anomaly detection in time series [C]// Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium: [s. n.], 2015: 89-94.
- [17] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization [EB/OL]. (2015-02-19) [2020-09-11]. <https://arxiv.org/pdf/1409.2329.pdf>.
- [18] MNH V, HEES N, GRAVES A, et al. Recurrent models of visual attention [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2204-2212.
- [19] XU K, WU L F, WANG Z G, et al. Exploiting rich syntactic information for semantic parsing with graph-to-sequence model [C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2018: 918-924.
- [20] CHEN B, SUN L, HAN X P. Sequence-to-action: end-to-end semantic graph generation for semantic parsing [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2018: 766-777.
- [21] SUN Y B, TANG D Y, DUAN N, et al. Semantic parsing with syntax- and table-aware SQL generation [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2018: 361-372.
- [22] DONG L, LAPATA M. Coarse-to-fine decoding for neural semantic parsing [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2018: 731-742.
- [23] YIN P C, NEUBIG G. TRANX: a transition-based neural abstract syntax parser for semantic parsing and code generation [C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Stroudsburg, PA: Association for Computational Linguistics, 2018: 7-12.
- [24] ZHANG X, HE S Z, LIU K, et al. AdaNSP: uncertainty-driven adaptive decoding in neural semantic parsing [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2019: 4265-4270.
- [25] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2017: 6000-6010.
- [26] SUN Z Y, ZHU Q H, XIONG Y F, et al. TreeGen: a tree-based transformer architecture for code generation [C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2020: 8984-8991.
- [27] DOZAT T, MANNING C D. Simpler but more accurate semantic dependency parsing [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2018: 484-490.
- [28] DOZAT T, MANNING C D. Deep biaffine attention for neural dependency parsing [EB/OL]. (2017-03-10) [2020-09-12]. <https://arxiv.org/pdf/1611.01734.pdf>.
- [29] WANG Y R, WU L J, XIA Y C, et al. Transductive ensemble learning for neural machine translation [C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2020: 6291-6298.
- [30] MANNING C D, SURDEANU M, BAUER J, et al. The Stanford CoreNLP natural language processing toolkit [C]// Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Stroudsburg, PA: Association for Computational Linguistics, 2014: 55-60.
- [31] ZETTLEMOYER L, COLLINS M. Online learning of relaxed CCG grammars for parsing to logical form [C]// Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Stroudsburg, PA: Association for Computational Linguistics, 2007: 678-687.
- [32] KWIATKOWSKI T, ZETTLEMOYER L, GOLDWATER S, et al. Lexical generalization in CCG grammar induction for semantic parsing [C]// Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2011: 1512-1523.
- [33] LIANG P, JORDAN M I, KLEIN D. Learning dependency-based compositional semantics[J]. Computational Linguistics, 2013, 29 (2): 389-446.
- [34] KWIATKOWSKI T, CHOI E, ARTZI Y, et al. Scaling semantic parsers with on-the-fly ontology matching [C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2013: 1545-1556.
- [35] ZHAO K, HUANG L. Type-driven incremental semantic parsing with polymorphism [C]// Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2015: 1416-1421.
- [36] JIA R, LIANG P. Data recombination for neural semantic parsing [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2016: 12-22.
- This work is partially supported by the Major Program for Technological Innovation 2030 - "New Generation Artificial Intelligence" (2018AAA0100500), the Guangdong Province Science and Technology Program (2017B010110011).
- XIE Defeng**, born in 1994, M. S. candidate. His research interests include semantic analysis, machine reading comprehension, named entity recognition.
- Ji Jianmin**, born in 1984, Ph. D., associate professor. His research interests include natural language processing, deep reinforcement learning, cognitive robotics.