



计算机应用
Journal of Computer Applications
ISSN 1001-9081, CN 51-1307/TP

《计算机应用》网络首发论文

题目：融入句法感知表示进行句法增强的语义解析
作者：谢德峰，吉建民
收稿日期：2020-11-27
网络首发日期：2021-02-07
引用格式：谢德峰，吉建民. 融入句法感知表示进行句法增强的语义解析[J/OL]. 计算机应用. <https://kns.cnki.net/kcms/detail/51.1307.tp.20210205.1445.015.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

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 230026)

(* 通信作者电子邮箱 jianmin@ustc.edu.cn)

摘要:在自然语言处理中,句法信息是完整句子中字词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、计算机可执行的意义表示。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度,提出一种利用输入端句法依存关系信息来增强模型效率的方法。该方法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,而后将其可看作为是句法感知表示的中间表示用作为语义解析模型部分输入嵌入,并采用转导融合学习方式对模型融合。实验对比了基准模型Transformer以及过去十年的相关工作。最终实验结果表明,在ATIS, GEO, JOBS数据集上,该句法增强语义解析方法分别实现了89.1%, 90.7%, 91.4%的最佳准确率,不仅全面超过了Transformer,还在大部分指标上超越了其他以前的相关工作。这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 230026, China);

(* The corresponding author: jianmin@ustc.edu.cn)

Abstract: Syntactic information, which implies syntactic structure relations or dependency relations between words of a complete sentence, is an important and effective reference in natural language processing. The task of semantic parsing is to transform natural language statements directly into semantic complete and computer-executable formal representations. In previous semantic analysis studies, there were few efforts on improving the efficiency of end-to-end semantic parsing by using syntactic information from input sources. To further improve the accuracy of an end-to-end semantic parsing model, a novel method was proposed to integrate source-side dependency relation information implicitly for semantic parsing. The basic idea is to pre-train an end-to-end dependency parser firstly, then exploit the intermediate hidden representations, i.e., syntax-aware representation, as parts of input embedding to the semantic parser, where models were fused by transductive ensemble learning in experiments. A Transformer model based end-to-end semantic parser was used as the baseline for comparison. Experimental results show that, on ATIS, GEO and JOBS datasets, the method respectively achieve the best accuracy of 89.1%, 90.7%, and 91.4%, which not only exceed the performance of the baseline, but also surpasses previous state-of-the-art on most of indicators. It verify the effectiveness of integrating dependency relation information.

Key words: syntax-aware; semantic parsing; deep learning; natural language processing; 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,成为了二十一世纪最火热的话题。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]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)的定义就是这样一个任务: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

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,中国科学技术大学副教授,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。

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。如何在上述的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,成为了一个急需研究和解决的难题。

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)。请注意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(包括结构关系或依存关系),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、较为复杂的一种任务。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。句法结构树的层级关系如何自上而下分解原句的信息,依存关系树则描述了原句中词与词之间的依存关系。

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。综上,为了进一步提高语义解析模型的准确度和效率,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单元(tree long short-term memory, Tree-LSTM)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。本文工作受文献[7]的思路启发,但与文献[7]

中的机器翻译工作相比,本文的任务目标,所采用依存关系解析器模型,语义解析基础模型,实验方法步骤,和训练数据集都与文献[7]的工作不同。同时语义解析和机器翻译作为两个不同的研究问题有较大差别。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五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。

本文的贡献和创新点在于:首次将融入句法感知表示的思路用于语义解析任务,首次用句法感知表示的方法改进基于Transformer的语义解析模型。改进旨在利用句法信息来增强模型效率。与此同时,本文工作也是首次在语义解析任务中采用转导融合学习来进行模型融合。

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,也很难扩展到通用的领域。

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][9][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的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解决语义解析任务,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,以此提升组合范畴文法的效率。

随着深度学习的发展,许多基于神经网络seq2seq模型或其衍生模型呼之欲出。深度学习的优点在于其能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]中采用基于长短期记忆单元(Long short-term memory, LSTM)^{[16][17]}和基于注意力机制^{[3][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问题中的结构建模问题,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]则相反,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 reader, 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,然后实现了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,Transformer 在处理大规模数据集上比 LSTM 要更具鲁棒性,不容易过拟合。文献[21]的研究团队后续进一步研究提出的 TreeGen^[26]就是借助了 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。Transformer 在语义解析上的应用探究将会是未来几年内的一个趋势和方向。

本文将在第2章节探讨融入依存关系信息表示的句法增强模型所用到的模型和方法,在第3章节说明为证明该模型的有效性而展开的若干实验及实验细节。

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可以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。

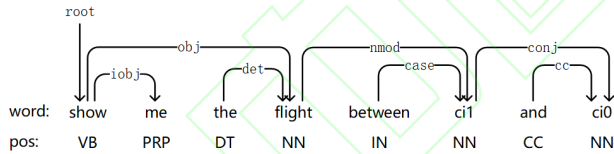


图1 依存关系信息示例

Fig. 1 An example of dependency information

依存关系解析器的作用是将输入句子及该句子的词性标注,在句子的基础上将每个词在句子中的依存关系标记出来,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。

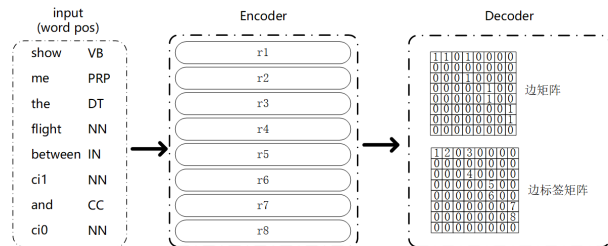


图2 依存关系解析的解析过程示意

Fig. 2 The process of dependency parsing

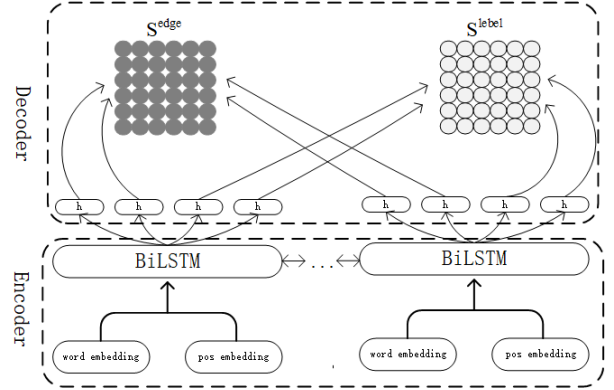


图3 依存关系解析器模型结构示意图

Fig. 3 The structure of the dependency parser

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,因此我们采用该依存关系解析器用于提取输入句子中的依存关系信息。其结构如图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,如图3所示,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),将输入的句子的词嵌入(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 long short-term memory, BiLSTM)处理该输入,得到中间表示 R 。如公式(1)和公式(2),其中 $e_i^{(word-dependency)}$, $e_i^{(pos)}$ 为依存关系解析器端的词嵌入和标注信息嵌入。

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

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

$$\text{Bilin}(x_1, x_2) = x_1^T U x_2 \quad (3)$$

$$\text{Biaff}(x_1, x_2) = x_1^T U x_2 + W(x_1 \oplus x_2) + b \quad (4)$$

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

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

$$h_i^{(edge-dep)} = \text{FNN}^{(edge-dep)}(r_i) \quad (7)$$

$$h_i^{(label-head)} = \text{FNN}^{(label-head)}(r_i) \quad (8)$$

$$s_{i,j}^{(edge)} = \text{Biaff}^{(edge)}(h_i^{(edge-dep)}, h_j^{(edge-head)}) \quad (9)$$

$$s_{i,j}^{(label)} = \text{Biaff}^{(label)}(h_i^{(label-dep)}, h_j^{(label-head)}) \quad (10)$$

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

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

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

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

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

随后,用四个不同的全连接层分别处理中间表示 R ,得到四个表示 $h_i^{(edge-head)}$, $h_i^{(edge-dep)}$, $h_i^{(label-head)}$, $h_i^{(label-dep)}$,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,达到防止过拟合的作用。

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],如公式(9)(10)所示,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 $s^{(edge)}$ 时, U 的维度为 $(d \times 1 \times d)$, d 表示 $h_i^{(edge-head)}$, $h_i^{(edge-dep)}$, $h_i^{(label-head)}$, $h_i^{(label-dep)}$ 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 $s^{(label)}$ 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 $s^{(edge)}$ 边矩阵中, $s^{(edge)}[i][j]$ 表示节点 i 指向节点 j 的边存在的概率,而 $s^{(label)}$ 边标签矩阵中, $s^{(label)}[i][j]$ 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。

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。过程如公式(13)(14)(15)所示。

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。如公式(16)所示。

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

decoder部分保持与Transformer的decoder部分一致。如图4所示,图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。可以看到,原本的Transformer只需要输入词嵌入+位置编码,采用本文所提方法改进后,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。

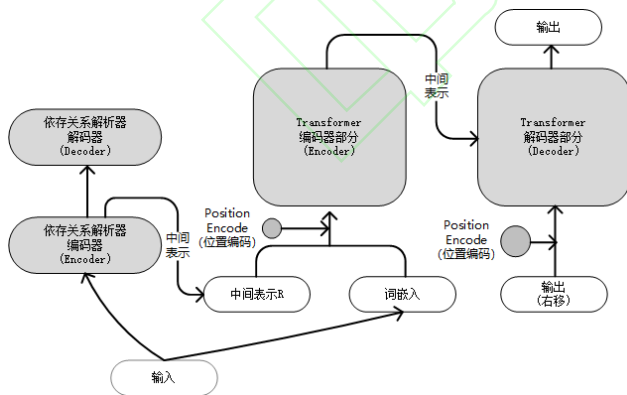


图4 融入句法信息感知后的模型整体结构示意图

Fig. 4 Structure of the Transformer model exploiting 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,本文首次将转导融合学习用于语义解

$$\text{sub_layer_output} = \text{LayerNorm}(\text{sub_layer_input} + (\text{SubLayer}(\text{sub_layer_input}))) \quad (16)$$

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

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

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

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)。上述三个步骤如公式(17)(18)(19)所示。

decoder部分跟encoder相比,多了一个层用于处理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,也就是公式(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:

$$(20)$$

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,转导融合学习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。相比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)、(22)、(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(23)中的任意一个子模型,用合成数据集按对其进行微调,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。

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

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

$$f_1, f_2, f_3, f_4, \dots, f_K \quad (23)$$

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

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

3 实验和实验结果分析

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 about the datasets

数据集	ATIS	GEO	JOBS
Train	4434	600	500
Dev	491	-	-
Test	448	280	140
平均输入长度	10.5	7.4	8.7
最大输入长度	44	23	22
平均标注长度	28.2	19.1	10.0
最大标注长度	94	51	51

3.2 实验设置和细节

3.2.1 评价指标

对于 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结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[15]中的度量标准,文献[15]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 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 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 (earlystop) 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。本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 和表 3 所示。表中,DR 表示的是模型该处的丢弃率 (dropout rate, DR), DI 表示的是该处的维度 (dimension, DI)。

3.3 实验结果

本课题主要在不同数据集上做了四组实验,实验 1 为基

表 2 依存关系解析器的参数设置

Tab. 2 Parameters of the dependency 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

于 Transformer 的语义解析模型作为实验的基准模型 (baseline), 实验 2 是在实验 1 的基础上加入了转导融合学习, 实验 3 为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 则是在实验 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本实验主要模型, 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, 表 4 中前六个实验数据陈列了过去十年的传统机器学习方法的实验结果, 第七到第十五个实验数据则是陈列了近几年基于神经网络的方法的实验结果, 最后是本文四个实验的实验结果。JOBS 数据集上的实验结果如表 5 所示, 表 5 中前三个实验数据陈列了过去十年的传统机器学习方法的实验结果, 第四到第六个实验数据则是陈列了近几年基于神经网络的方法的实验结果, 最后是本文四个实验的实验结果。表 4、表 5 中用 * 标明了两个使用外延匹配 (Denotation Match) 方式进行结果评估的实验对比数据。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。

表 3 语义解析器的参数设置

Tab. 3 Parameter of the semantic parser

参数	值	参数	值
word embedding DI	512	Feedforward DI	2048
encoder layer	2	decoder layer	3
head number	8	attention dim DI	512
activation	relu	dropout rate	0.1
batch size	16	learning rate	1e-4

表 4 ATIS, GEO 数据集上的实验结果

Tab. 4 Experimental results on ATIS and GEO

算法模型	ATIS		GEO	
	Exact match	Tree match	Exact match	Tree match
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	-
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
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

3.4 实验结果分析

对比实验 1 和以往基于 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都要比不融入该信息的模型表现要更好, 证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, 分析得到可能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 指标上表现良好却反而在 Exact match 指标上表现一般。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。这证明了引入句法依存关系信息的有效性。

表 5 JOBS 数据集上的实验结果

Tab. 5 Experimental results on JOBS

算法模型	JOBS
	Exact match
ZC07 ^[31]	79.3
DCS ^[33]	90.7
TISP ^[35]	85.0
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. (JI J M. Several methods to improve ASP efficiency and application on service robot [D]. Hefei: University of Science and Technology of China, 2010.)
- [2] 柴伟. 短语结构句法分析综述 [J]. 电脑知识与技术, 2020, 16 (16): 26-30. (CAI W. Summary of phrase structure syntactic analysis [J]. Computer Knowledge and Technology, 2020, 16 (16): 26-30.)
- [3] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]// ICLR 2015: Proceedings of the 3rd International Conference on Learning Representations. 2015.
- [4] CHO K, VAN MERRIENBOER B, GÜLÇEHRE Ç, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]// EMNLP 2014: 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 (Volume 1: Long Papers). 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, LI Z, FU G, 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, Volume 1 (Long and Short Papers). 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, Cambridge, MA: MIT Press, 2014: 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, MIT Press, 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 2012 International Conference on Machine Learning. New York: ACM, 2012.
- [13] ZETTLEMOYER L S, COLLINS M. Learning to map sentences to logical form: structured classification with probabilistic categorical grammars [C]// Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence. San, Francisco: AUAI, 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. 2015: 89-94.
- [17] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization [J]. arXiv preprint arXiv: 1409.2329. 2014.
- [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-Volume 2. Cambridge, MA: MIT press, 2014: 2204-2212.
- [19] XU K, WU L, WANG Z, 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. 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 (Volume 1: Long Papers). Stroudsburg, PA: Association for Computational Linguistics, 2018: 766-777.
- [21] SUN Y, TANG D, 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 (Volume 1: Long Papers). 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 (Volume 1: Long Papers). Stroudsburg, PA: Association for Computational Linguistics, 2018: 731-742.
- [23] YIN P, NEUBIG G. TRANX: A transition-based neural abstract syntax parser for semantic parsing and code generation [C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing (Demo Track). Stroudsburg, PA: Association for Computational Linguistics, 2018: 7-12.
- [24] ZHANG X, HE S, 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]// Advances in Neural Information Processing Systems. Cambridge, MA: MIT press, 2017: 5998-6008.
- [26] SUN Z, ZHU Q, XIONG Y, et al. TreeGen: a tree-based transformer architecture for code generation [C]// AAAI: The Thirty-Fourth AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI, 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.
- [28] DOZAT T, MANNING C D. Deep biaffine attention for neural dependency parsing [J]. arXiv preprint arXiv: 1611.01734. 2016.
- [29] WANG Y, WU L, XIA Y, et al. Transductive ensemble learning for neural machine translation [C]// AAAI: The Thirty-Fourth AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI, 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] ZETTMLOYER L, COLLINS M. Online learning of relaxed CCG grammars for parsing to logical form [C]// EMNLP-CoNLL: 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, ZETTMLOYER 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]// HLT-NAACL: Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2015.
- [36] JIA R, LIANG P. Data recombination for neural semantic parsing [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: Association for Computational Linguistics, 2016: 12-22.

This work is partially supported by the National Major Program for Technological Innovation 2030—New Generation Artificial Intelligence (2018AA000500), the Guangdong Province Science and Technology Plan project (2017B010110011).

XIE Defeng, born in 1994, M. S. candidate. His research interests include semantic analysis, reading comprehension, named entity recognition.

Ji Jianmin, born in 1984, Ph. D., associate professor. His research interests include natural language processing, deep reinforcement learning, and cognitive robotics