

基于 BiLSTM 并结合自注意力机制和句法信息的 隐式篇章关系分类

凡子威 张 民 李正华

(苏州大学计算机科学与技术学院 江苏 苏州 215006)

摘 要 隐式篇章关系分类是浅层篇章结构分析(Shallow Discourse Parsing)中的子任务,也是自然语言处理(Natural Language Processing, NLP)中的一项重要任务。隐式篇章关系是由篇章关系中的论元对推理出来的逻辑语义关系。隐式篇章关系的分析结果可以应用于许多自然语言处理任务中,如机器翻译、自动文档摘要、问答系统等。针对隐式篇章关系分类任务,提出一种基于自注意力机制和句法信息的方法。通过双向长短时记忆网络(Bidirectional Long Short-Term Memory Network)对输入的结合句法信息的论元对进行建模,将论元对表示成低维稠密的向量;通过自注意力机制对论元对信息进行筛选。在 PDTB2.0 数据集上进行实验,结果表明该方法较基准系统获得了更好的效果。

关键词 神经网络,隐式篇章关系分类,自注意力机制,句法信息

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.05.033

BiLSTM-based Implicit Discourse Relation Classification Combining Self-attention Mechanism and Syntactic Information

FAN Zi-wei ZHANG Min LI Zheng-hua

(School of Computer Sciences and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract Implicit discourse relation classification is a sub-task in shallow discourse parsing, and it's also an important task in natural language processing(NLP). Implicit discourse relation is a logic semantic relation inferred from the argument pairs in discourse relations. The analytical results of the implicit discourse relationship can be applied to many natural language processing tasks, such as machine translation, automatic document summarization, and question answering system. This paper proposed a method based on self-attention mechanism and syntactic information for the classification task of implicit discourse relations. In this method, Bidirectional Long Short-Term Memory Network (BiLSTM) is used to model the inputted argument pairs with syntactic information and express the argument pairs into low-dimension dense vectors. The argument pair information was screened by the self-attention mechanism. At last, this paper conducted experiments on PDTB2.0 dataset. The experimental results show that the proposed model achieves better effects than the baseline system.

Keywords Neural network, Implicit discourse relation classification, Self-attention mechanism, Syntactic information

1 引言

篇章作为词和句子更上层的一种分析粒度,在自然语言理解和自然语言生成中起到至关重要的作用,与之相对应的浅层篇章结构分析是自然语言处理中一项具有重大意义的任务,它是自然语言理解的基础。近年来,篇章理论的发展以及大规模篇章语料的发布使得浅层篇章结构分析系统的构建越发重要。篇章不是简单的由句子堆砌成的序列,而是由一系列语义连贯的短语、子句、句子或者段落构成的具有独立语义

的自然语言文本。篇章内部的论元之间的关系可以实现篇章文本的结构化和层次化,为篇章的结构和语义分析提供坚实的理论和研究基础。篇章分析的结果具有广泛的应用价值,例如:在统计机器翻译(Statistics Machine Translation, SMT)中,通过识别文本中的篇章连接词,来提升统计机器翻译的性能^[1];在问答系统(Question Answering, QA)中,通过将篇章信息整合到问题和答案对的相似度计算中,使得模型的性能有极大的提升^[2];在自动摘要系统中,通过识别篇章中的因果、转折等篇章关系来选取文章中具有重要意义的句子作为

到稿日期:2018-04-01 返修日期:2018-07-29 本文受国家自然科学基金项目(61525205,61876116)资助。

凡子威(1992—),男,硕士生,主要研究方向为自然语言处理,E-mail:1204493155@qq.com;张 民(1970—),男,博士,教授,主要研究方向为自然语言处理、机器翻译、人工智能;李正华(1983—),男,博士,副教授,主要研究方向为自然语言处理、句法分析、中文分词、语义分析、篇章分析,E-mail:zhli13@suda.edu.cn(通信作者)。

摘要的备选句子^[3]。

基于 PDTB 语料,研究人员将浅层篇章分析研究任务细化为篇章连接词识别、浅层篇章结构关系的论元对识别、论元对间显式篇章关系分类和论元对间隐式篇章关系分类这 4 个子任务。浅层篇章结构中的显式篇章关系这个语言现象中存在连接词这一具有启发意义的短语片段,连接词的准确识别是判定与该连接词相关联的两个论元对之间的显式篇章关系的基础。连接词本身在篇章中是具有歧义性的,需要通过分类的方法去识别该连接词是否起到连接两个论元的作用。Pitler 等^[4]指出篇章连接词的歧义性主要体现在以下两个方面:1)该词在篇章中是否起连接作用;2)该词在表示具体连接关系类型时是否存在歧义。相对于连接词识别任务,论元对的识别任务具有很大的难度,论元对之间的相对位置的确定是篇章分析研究的难点,目前研究者倾向于用序列标注的方法来识别论元 1 和论元 2 在句子中的范围。隐式篇章关系由于缺少连接词这一直接线索,只能依据论元对从句法、语义、上下文中抽取相关特征进行分析判断。上下文信息的不确定性、句子结构的复杂性和语义关系的歧义性将影响隐式篇章关系的推理过程。使用人工提取特征的方法不能有效地提高隐式篇章关系分类任务的性能。随着深度学习研究的飞速发展,人工难以提取有效的特征的难题得到了解决,即将篇章中的词和句子映射成低维的稠密向量,用这些向量来表示篇章。通过向量表示的论元对可以在高维空间中体现论元对之间的相关性,从而提取到有用的特征。下面展示一个具体的隐式篇章关系的例子:

[*The Arabs had merely oil.*]_{Arg1} Implicit=while [**These farmers may have a grip on the world's very heart.**]_{Arg2} (Comparison, Contrast)[wsj 0515]

其中,斜体部分代表的是论元 1,加粗部分代表的是论元 2,“Comparison, Contrast”为隐式篇章关系类型。上例中含有下划线的连接词“while”是标注者推断出的连接词,在原文中该处是不含有连接词的。研究者可以借助标注者推断的连接词推测出论元对所蕴含的隐式篇章关系类别,但是不可以将该连接词作为真实的篇章连接词使用。

本文主要针对隐式篇章关系分类任务展开研究,在 BiLSTM 网络的基础上探索了结合自注意力机制和添加句法信息的模型,构建了新的隐式篇章关系分类系统。本文方法不需要通过构建人工特征的方式对论元对进行建模,其通过将论元对表示成词向量的形式可以有效地避免人工特征选择的过程,改善了传统方法中使用人工特征建模论元对产生的主观随机性的问题,因而可以方便快捷地构建隐式篇章关系分类系统。在 PDTB 数据上的实验结果表明,使用自注意力机制并结合句法信息的隐式篇章关系分类模型能够获得比基准模型(BiLSTM)更好的效果。

2 相关工作

近年来,浅层篇章结构分析任务得到了很多研究者的关注。2008 年发布了第一个人工标注的英文篇章关系语料库

Penn Discourse TreeBank 2.0(PDTB2.0)^[5],其公开之后产生了大量与浅层篇章结构分析相关的工作。浅层篇章结构分析主要包括 4 个子任务:1)篇章连接词识别;2)浅层篇章结构关系的论元对识别;3)显式篇章关系识别;4)隐式篇章关系识别。其中浅层篇章结构关系相关的论元对识别又可以拆分成显式篇章关系的论元对识别和隐式篇章关系的论元对识别。本文主要针对隐式篇章关系分类任务展开研究。目前,学术界通常采用隐马尔科夫模型(Hidden Markov Model, HMM)^[6]、最大熵模型^[7]、感知器模型^[8]、支持向量机^[9]、条件随机场模型(Conditional Random Field, CRF)^[10]、前馈神经网络、循环神经网络(Recurrent Neural Networks, RNN)、长短时记忆网络(Long Short Term Memory Networks, LSTM)、卷积神经网络(Convolutional Neural Network, CNN)等方法来处理隐式篇章关系分类任务。

隐式篇章关系分类任务不同于显式篇章关系分类任务,显式篇章关系有连接词作为提示信息指导论元对之间蕴含的逻辑语义关系,而隐式篇章关系只有论元对作为输入信息,缺少连接词这一重要的指示信息。研究者需要根据给定的论元对来识别所蕴含的隐式篇章关系类型。隐式篇章关系分类的性能目前普遍较低,研究者对其进行深入研究以提高该任务的性能。Pitler 等^[11]首先提出了在 PDTB 数据集上针对隐式篇章关系进行探究,针对隐式篇章关系的论元对构建了相关的词法、句法特征,并引入了情感特性特征,结合最大熵模型辅助推理隐式篇章关系类型。Lin 等^[12]结合论元对的成分句法树和依存句法树结构信息提出了相关的句法特征,使得模型性能有了一定的提升。Wang 等^[13]提出了基于树核的方法自动地从句法树中抽取相应的特征,提高了隐式篇章关系分类模型的性能。Rutherford 等^[14]提出了使用 Brown 聚类词对特征来提高隐式篇章关系模型的性能。Qin 等^[15]提出了一个新颖的方法,即使用卷积神经网络模型构建隐式篇章关系分类器,实现隐式篇章关系的分类。该卷积神经网络的架构使用了一层卷积层和一层 maxpooling 层,之后连接了一层 softmax 层。在构建卷积神经网络的输入层的输入向量时,分别将论元 1 和论元 2 的单词和单词的词性映射成多维向量,并将单词向量和词性向量进行拼接,使输入向量带有词性的特征;然后将分别经过卷积之后的代表论元 1 和论元 2 的向量进行拼接,输入到 softmax 层预测模型的输出。在开发集上,分类器在给定标准的论元 1 和论元 2 数据后,对隐式关系分类的 F1 值可以达到 46.33%。Schenk 等^[16]提出了一个新颖的方法,即使用含有一层隐藏层的前馈神经网络构建隐式篇章关系分类器,实现对隐式篇章关系的分类。在构建输入层的输入向量时,将论元 1 所在句子的词分别映射成词向量,将这组词向量相加后取平均得到的向量和这组词向量对应位置相乘后得到的词向量相加,得到一个可以表示论元 1 的句子向量,论元 2 与论元 1 的处理相似,将得到的两个句子向量进行拼接得到输入层的输入向量。应用这种结构的分类器能够得到一个较好的分类结果,在开发集上分类器在给定标准的论元 1 和论元 2 数据后,对隐式关系的分类的 F1 值

达到 45.42%。Weiss 等^[17]应用循环神经网络实现了隐式篇章关系分类,提出了一种简洁的框架,将论元 1、论元 2、连接词、标点符号映射成稠密的向量,分别输入到 4 个对应的循环神经网络中,将 4 个循环神经网络的输出向量拼接成一个向量输入到含有一层隐藏层的全连接层中,经过 softmax 层输出预测结果。该模型达到了一个很好的分类效果,在开发集上对隐式关系识别的 F1 值达到了 34.58%。Chen 等^[18]对隐式篇章关系中的顶层类别 Comparison、Contingency、Expansion、Temporal 进行分类,将 EntRel 关系实例作为 Expansion 关系实例添加到 Expansion 中,扩充了 Expansion 的实例数目。其使用 BiLSTM 和 GRU 模型构建浅层篇章结构隐式篇章关系模型,其中 Bi-LSTM 可以很好地获取当前词的上下文信息,GRN 可以有效地解决两个句子之间的语义鸿沟问题。

3 语料资源

3.1 PDTB2.0

Penn Discourse Treebank 2.0^[4]是目前已公开发布的最大的人工标注的篇章关系树库,它是针对篇章的浅层结构进行关系标注。篇章关系树库在华尔街时报(Wall Street Journal, WSJ)的文章上展开标注,共包含 2312 篇文章,文章被分为 25 个章节,为 Sections 00-24。PDTB2.0 中给出了 5 种浅层篇章结构关系类型,分别是显式篇章关系(Explicit)、隐式篇章关系(Implicit)、实体型关系(EntRel)、可替代关系(AltLex)、空关系(NoRel)。表 1 列出了 PDTB2.0 数据集中这 5 种关系的分布情况。

表 1 PDTB2.0 浅层篇章关系数据的分布

Table 1 Statistics of shallow discourse relation in PDTB2.0

浅层关系类型	数量
Explicit	18459
Implicit	16224
EntRel	5210
AltLex	624
NoRel	254

PDTB2.0 中将浅层篇章结构关系类型分为 3 层级联结构。第 1 层为 Class 层,主要包含 4 种第一级浅层篇章结构关系类型,分别为 Comparison (Comp.), Contingency (Cont.), Expansion (Exp.), Temporal (Temp.)。第 2 层为 Type 层,是 Class 层的子层,包含了 16 种关系类别。第 3 层为 Subtype 层,为 Type 层的进一步细化。表 2 列出了 PDTB2.0 中的浅层篇章结构关系类型的前两层(Class 和 Type)的具体语义关系类型。

表 2 PDTB2.0 Class 和 Type 级关系类型的数据分布

Table 2 Statistics of Class and Type relation in PDTB2.0

Class	Type
Comparison	Contrast, Pragmatic Contrast, Concession, Pragmatic Concession
Contingency	Cause, Pragmatic Cause, Condition, Pragmatic Condition
Expansion	Conjunction, Instantiation, Restatement, Alternative, Exception, List
Temporal	Asynchronous, Synchronous

本文针对 PDTB2.0 数据集中的 Class 类别展开研究工作。表 3 列出了 Class 类别中的 4 个浅层篇章结构关系类型在训练集、开发集、测试集中的数据分布情况。

表 3 PDTB2.0 隐式关系类型 Class 层的数据分布

Table 3 Statistics of implicit discourse relation Class in PDTB2.0

浅层关系类型	训练集	开发集	测试集
Comparison	1944	197	152
Contingency	3346	292	279
Expansion	7011	671	574
Temporal	760	64	85

4 基于 BiLSTM 自注意力机制和句法信息的隐式篇章关系分类

4.1 基于 BiLSTM 网络的隐式篇章关系分类

本节使用 BiLSTM 网络构建了隐式篇章关系分类的基准模型,介绍了数据对应的标签集合和基于 BiLSTM 网络的隐式篇章关系分类模型的实现过程。

4.1.1 标签集合

我们对 PDTB 数据集中的 Class 类型展开分析研究,以下的模型都是建立在 Class 层级的基础上,分别为 Comparison, Contingency, Expansion, Temporal。

4.1.2 BiLSTM 模型

BiLSTM (Bi-directional Long Short-Term Memory Network) 是在 LSTM (Long Short-Term Memory Network, LSTM) 网络上的拓展,采用了双向的 LSTM 的机制对句子进行建模,提取出相对于 LSTM 网络更丰富的句子信息。LSTM 网络使用了输入门、遗忘门、输出门机制来控制 LSTM 内部记忆单元对之前的信息的记忆程度。一个完整的 LSTM 的 cell 可由式(1)表示:

$$\begin{aligned}
 f_t &= \sigma(W_f[h_{t-1}, e^{w_t}] + b_f) \\
 i_t &= \sigma(W_i[h_{t-1}, e^{w_t}] + b_i) \\
 o_t &= \sigma(W_o[h_{t-1}, e^{w_t}] + b_o) \\
 \tilde{C}_t &= \tanh(W_c[h_{t-1}, e^{w_t}] + b_c) \\
 C_t &= \tilde{C}_t * i_t + C_{t-1} * f_t \\
 h_t &= o_t * \tanh(C_t)
 \end{aligned} \tag{1}$$

其中, σ 代表 sigmoid 激活函数; i, f, o 和 c 分别代表输入门、遗忘门、输出门和 LSTM 的隐藏信息,它们和 h_t 具有相同的维度大小; e^{w_t} 是输入的句子中词 w_t 映射产生的词向量, $e^{w_t} \in R^{d_c}$; d_c 是词向量的维度。本文实现了将基于 BiLSTM 网络的隐式篇章关系分类系统作为基准系统,图 1 展示了基于 BiLSTM 的隐式篇章关系分类模型,其中的 LSTM 单元代表了式(1)所示的 LSTM cell。

模型的输入层是论元 1 和论元 2 的句子映射成的词向量,论元 1 中的词为 $X^1 = [w_1, w_2, \dots, w_n]$ (n 为论元 1 的句子长度),论元 2 中的词为 $X^2 = [w_1, w_2, \dots, w_m]$ (m 为论元 2 的句子长度)。首先对论元 1 和论元 2 中的每一个词通过词向量映射的形式转化为词向量表示,得到式(2)所示的两个句子的向量矩阵。其中 d_c 是词向量的维度。

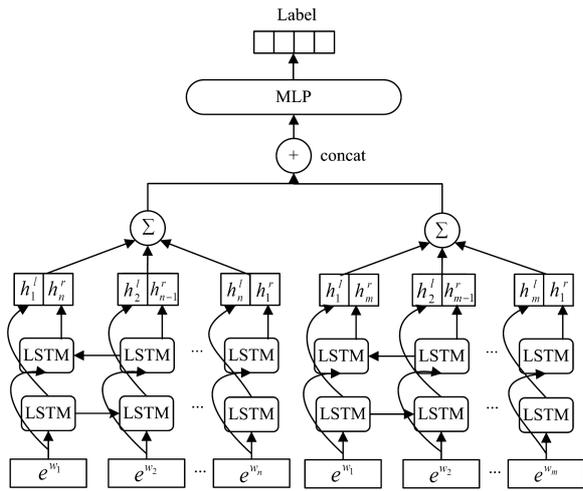


图 1 基于 BiLSTM 网络的模型
Fig. 1 Model based on BiLSTM

$$e_{\text{arg1}}^w = [e^{w_1}, e^{w_2}, \dots, e^{w_n}], e^{w_i} \in R^{d_e}, i \in [1, n]$$

$$e_{\text{arg2}}^w = [e^{w_1}, e^{w_2}, \dots, e^{w_m}], e^{w_j} \in R^{d_e}, j \in [1, m]$$

如图 1 所示,输入层的论元 1 和论元 2 的向量表示输入到正向的 LSTM 层和反向的 LSTM 层即 BiLSTM 层,通过 BiLSTM 层可以得到论元 1 和论元 2 的抽象表示,如式(3)所示:

$$h_{\text{arg1}} = \sum_{i=1}^n h_i^l \oplus h_r^{n-i+1}$$

$$h_{\text{arg2}} = \sum_{j=1}^m h_j^l \oplus h_r^{m-j+1}$$

其中, h_i^l 为正向 LSTM 生成的相对于 e^{w_i} 的抽象表示, h_r^{n-i+1} 为反向 LSTM 生成的相对于 e^{w_i} 的抽象表示, h_{arg1} 和 h_{arg2} 是论元 1 和论元 2 经过 BiLSTM 的抽象表示。通过式(4)可以得到论元 1 和论元 2 的聚合表示:

$$h_{\text{args}} = h_{\text{arg1}} \oplus h_{\text{arg2}}$$

将式(4)得到的论元 1 和论元 2 的聚合向量作为 MLP (Multi-layer Perceptron) 的输入后经过 softmax 函数得到最终的相对于各个标签的分数,通过最高的分数得到对应的标签,如式(5)所示:

$$h_t = \tanh(W_t h_{\text{args}} + b_t)$$

$$p = \text{soft max}(W_y h_t + b_y)$$

$$y^* = \arg \max_{y \in Y} p$$

其中, Y 是标签集合,对应 PDTB 数据集集中的 Class 类别。

4.2 基于 BiLSTM 和自注意力机制的方法

自注意力机制 (Self-attention mechanism) 通常不会使用其他额外的信息,只关注句子本身,从句子中抽取相关信息。自注意力又称作内部注意力,本文在 BiLSTM 模型的基础上添加了自注意力机制,并使用了两种方式处理自注意力机制的输出。

1) 相加聚合,将自注意力模型编码论元对之后产生的隐藏向量按照对应的维度相加;

2) 向量拼接,将自注意力模型编码论元对之后产生的隐藏向量拼接。

结合自注意力机制的双向长短时记忆模型的输入同

BiLSTM,自注意力的关键部分如式(6)所示,将 BiLSTM 模型产生的隐藏层表示通过自注意力机制再次编码,提取出更高层次的特征表示。具体的模型结构如图 2 所示。

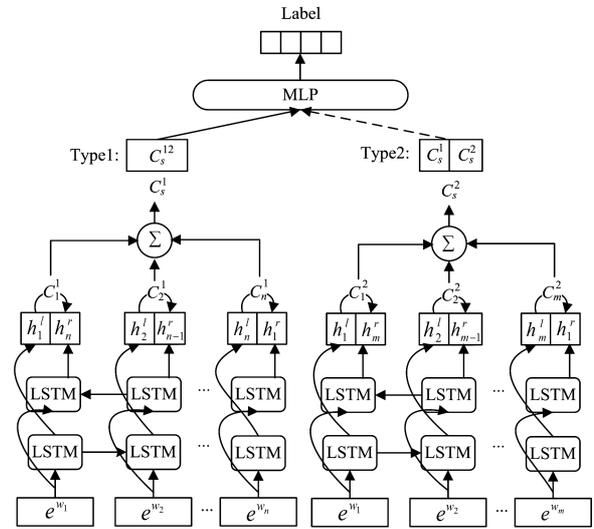


图 2 结合自注意力机制和 BiLSTM 网络的模型

Fig. 2 Model based on BiLSTM and self-attention mechanism

$$P_1 = \text{soft max}(V_1 \tanh(W_1 (h^1)^T))$$

$$C^1 = P_1 * h^1$$

$$P_2 = \text{soft max}(V_2 \tanh(W_2 (h^2)^T))$$

$$C^2 = P_2 * h^2$$

将 C^1 按照第一维度按位相加得到 C_s^1 ,将 C^2 按照第一维度按位相加得到 C_s^2 。式(7)展示了本文提出的两种自注意力机制生成隐藏向量的具体使用方法。

$$\text{Type1: } C_s^{12} = C_s^1 + C_s^2$$

$$\text{Type2: } C_s^{12} = C_s^1 \oplus C_s^2$$

4.3 基于句法信息的浅层篇章结构隐式关系分类

本文借鉴了传统机器学习方法处理隐式篇章关系分类任务时使用句法信息提高模型性能的思路,在上文中提出的基于自注意力机制和 BiLSTM 网络的隐式篇章关系分类模型的基础上将依存句法信息结合到任务中,通过给模型提供更高层次的句法输入信息,使得模型的性能有了明显的提升。本文创新性地使用了当前最好的预训练好的基于双仿射深度神经网络的依存句法分析模型^[19]的中间输出向量作为模型的依存句法信息的扩充输入。接下来将详细介绍基于双仿射深度神经网络的依存句法分析模型和基于句法信息的神经网络隐式篇章关系分类模型。

4.3.1 基于双仿射深度神经网络的依存句法分析模型

依存句法分析任务分为两个部分,第一部分是探究句子中的两个词之间是否具有相互依存的性质,另外一部分是探究具有依存性质的两个词之间具有何种依存关系。针对这两个任务,Dozat 等^[19]提出了基于双仿射深度神经网络的依存句法分析模型,该模型在公开数据集 PTB 上达到了目前最好的性能,模型的具体定义如下:

首先将训练集中的句子和句子对应的词性序列(其中词性标签序列是通过 Stanford CoreNLP 工具自动生成)分别映

射成向量的形式,句子中的每个词为 $X=[\omega_1, \omega_2, \dots, \omega_n]$ (n 为句子长度),句子对应的词性序列为 $T=[t_1, t_2, \dots, t_n]$ 。将句子中的词和词性转化为向量得到 e^w 和 e^t , 如式(8)所示:

$$\begin{aligned} e^w &= [e^{w_1}, e^{w_2}, \dots, e^{w_n}], e^{w_i} \in R^{d_e}, i \in [1, n] \\ e^t &= [e^{t_1}, e^{t_2}, \dots, e^{t_n}], e^{t_i} \in R^{d_t}, i \in [1, n] \end{aligned} \quad (8)$$

将句子通过映射转化后生成的词向量和词性向量拼接起来,得到双仿射神经网络依存句法模型的输入 e_i :

$$e_i = e^{w_i} \oplus e^{t_i}, i \in [1, n] \quad (9)$$

通过词向量和词性向量拼接的操作使得模型的输入既含有原始的词信息也含有更高层次的词性信息,让模型获得更充分的输入信息。

其次,通过使用多层 BiLSTM(Multi-layer BiLSTM)对输入的整个句子进行编码,可以获得当前词的表示 h_i ,该向量蕴含了当前词和当前词的上下文信息。

$$h_i = \text{BiLSTM}(e_i) \quad (10)$$

对于探究两个词之间是否具有依存性质的任务,以当前词语 i 为依存词,句子中除了当前词语 i 之外其他所有词为候选核心词,分别探究候选核心词与当前词 i 之间是否具有依存性质。为了防止 h_i 中含有过多的冗余信息,接下来使用了多层感知机(Multi-layer Perceptron, MLP)对向量进行降维操作,提炼出 h_i 中的重要信息,防止过多的冗余信息对模型产生干扰。对于句子中的每个词 h_i ,使用两个 MLP 分别得到这个词语作为核心词和依存词的表示,如式(11)所示:

$$\begin{aligned} r_i^{\text{arc-head}} &= \text{MLP}^{\text{arc-head}}(h_i), i \in [1, n] \\ r_j^{\text{arc-dep}} &= \text{MLP}^{\text{arc-dep}}(h_j), j \neq i, j \in [1, n] \end{aligned} \quad (11)$$

接下来通过双仿射(BiAffine)机制来计算当前依存词 i 和候选核心词之间的分数,如式(12)所示:

$$\text{score}(i \rightarrow j) = r_i^{\text{arc-head}} W r_j^{\text{arc-dep}} + r_i^h V, i, j \in [1, n] \quad (12)$$

4.3.2 基于句法信息的神经网络浅层篇章结构隐式关系分类模型

我们使用了 PTB 数据集作为训练数据,在训练好的双仿射依存句法模型的基础上,使用论元对作为依存句法模型的输入,获得了双仿射依存句法模型的 BiLSTM 层的输出信息。该信息是双仿射依存句法模型对论元对的抽象表示,通过预训练好的模型可以得到 BiLSTM 模型输出的含有依存句法信息的向量表示,具体的论元 1 和论元 2 的含有依存句法信息的 BiLSTM 输出如式(13)所示:

$$\begin{aligned} h^1 &= [h_1^1, h_2^1, \dots, h_n^1], h_i^1 \in R^{d_h}, i \in [1, n] \\ h^2 &= [h_1^2, h_2^2, \dots, h_m^2], h_j^2 \in R^{d_h}, j \in [1, m] \end{aligned} \quad (13)$$

通过上述方式获得了含有依存句法信息的论元对抽象表示 h^1 和 h^2 ,由此可得到结合句法信息的浅层篇章结构隐式关系分类模型的输入,如式(14)所示:

$$\begin{aligned} s_i^1 &= e_i^1 \oplus h_i^1, i \in [1, n] \\ s_j^2 &= e_j^1 \oplus h_j^1, j \in [1, m] \end{aligned} \quad (14)$$

基于 BiLSTM 的结合自注意力机制和句法信息的模型框架如图 3 所示。

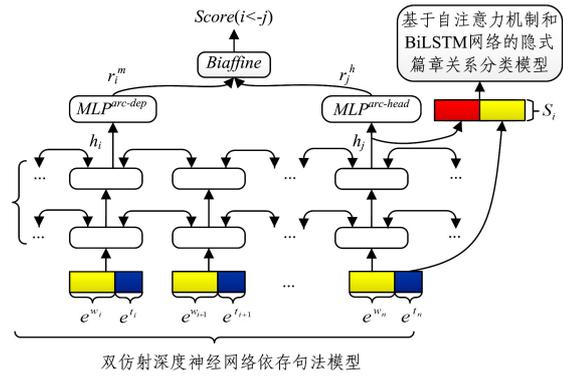


图 3 基于 BiLSTM 的结合自注意力机制和句法信息的模型

Fig. 3 BiLSTM-based model combining self-attention mechanism and syntactic information

5 实验

5.1 实验数据

本文在 PDTB2.0 数据集上进行了隐式篇章关系分类的实验,为了与前人的研究工作一致,本文使用了 Section 02-20 作为训练集,Sections 00-01 作为开发集,Section 21-22 作为测试集。PDTB2.0 数据在浅层篇章结构语义关系有 3 级结构,第 1 层为 Class,共分为 4 个语义类别,分别为 Comparison(Comp.), Contingency(Cont.), Expansion(Exp.), Temporal(Temp.)。第 2 层为 Type,共有 16 个类别,是在 Class 类别上的拓展。第 3 层为 Subtype,是在 Type 上的拓展。本文主要针对第 1 层 Class 的 4 个语义类别展开实验分析。对 PDTB2.0 数据集研究发现一个论元对实例有可能含有多个语义关系,数据的分布如表 4 所列。

表 4 PDTB2.0 论元对多语义数据分布

Table 4 Statistics of shallow discourse relation in PDTB2.0

语义关系数	训练集	开发集	测试集
1	12204	1142	1002
2	427	41	44
3	1	0	0
>3	0	0	0

如表 4 所列,我们在处理训练数据时,将含有多个浅层篇章结构关系的论元对拆分成多个训练实例,以保证每个论元对只含有一个对应的浅层篇章结构关系。在测试阶段模型预测出的浅层篇章结构关系类型如果在论元对对应的浅层篇章结构关系集合中,那么认为这次预测的结果是正确的。

5.2 实验设置

本文的基准系统是基于 BiLSTM 网络的隐式篇章关系分类模型,本文提到的所有模型都使用了 GoogleNews 数据集(约 1000 亿个词)上使用 word2vec 工具训练的 300 维的词向量。

基于 BiLSTM 网络的隐式篇章关系分类模型中,模型的超参数为 BiLSTM 结构中的隐藏层神经元的大小,我们尝试将隐藏层神经元数量设为 [50, 100, 200, 300],实验结果表明当神经元个数为 300 时效果最好,故 BiLSTM 模型最终选择了 300 维的隐藏层神经元个数。BiLSTM 模型中的全连接层使用了一层隐藏层,隐藏层神经元个数为 100。使用 tanh 作

为模型的激活函数。

基于 BiLSTM 和自注意力机制的隐式篇章关系分类模型中的权重矩阵 W_1 和 W_2 的维度为(600,600),权重向量 V_1 和 V_2 的维度为(1,600)。

基于句法信息的浅层篇章结构隐式关系分类模型中使用的基于双仿射深度神经网络的依存句法分析模型中的 LSTM 层的输出为 400 维。

本文所有使用神经网络的模型均使用交叉熵(Cross-Entropy)作为模型的损失函数,使用基于 AdaGrad 优化方法的批(mini-batch)梯度下降法训练模型的参数,批(Batch Size)的大小设置为 64,学习速率设置为 0.001。

5.3 实验结果和讨论

表 5 是基准模型和基于自注意力机制和句法信息的模型的开发集和测试集上的准确率。

表 5 模型在 PDTB 数据集上的结果
Table 5 Results of models on PDTB2.0

(单位:%)		
模型	开发集	测试集
BiLSTM(baseline)	56.94	56.88
Self-attention-Type1	56.58	58.12
Self-attention-Type2	57.11	58.79
Syntax+self-attention-type1	58.31	58.59
Syntax+self-attention-type2	57.37	58.98

表 5 的实验结果证明:添加自注意力机制的模型结果要好于 BiLSTM 模型,因为自注意力机制在 BiLSTM 模型的基础上通过注意力概率向量筛选了 BiLSTM 模型产生的特征,使得后续的网络在接收输入参数时能够降低噪音和冗余信息量,使得模型的效果更好。其中 Self-attention-Type2 模型在测试集上的性能比基准 BiLSTM 模型提高了 1.91%。表 6 是本文模型结果与前人模型结果的对比。

表 6 本文模型与其他模型的对比结果

Table 6 Comparison results of proposed model and other models
(单位:%)

模型	测试集
Zhang 等 ^[20]	55.39
Rutherford 等 ^[21]	57.10
Liu 等 ^[22]	57.57
Liu 等 ^[23]	57.27
Lan 等 ^[24]	58.95
本文模型	58.98

在隐式篇章关系分类任务上,相比只使用句子作为输入的神经网络方法,结合句法信息的神经网络方法取得了更好的性能。其原因是只使用句子信息作为输入的神经网络隐式篇章关系分类模型只能获得较为浅层的输入信息,模型只能根据输入的句子通过模型抽象地提取相关的特征信息,句法信息属于上层的辅助信息,无法通过简单的模型提炼出来。因为预训练好的双仿射依存句法模型的中间输出向量是拟合依存句法分析任务的,向量中包含了依存句法的信息,所以以其作为神经网络隐式篇章关系模型的输入能够为模型提供充分的输入信息。从实验结果可以看出: Syntax+self-attention-type1 的模型性能比 Self-attention-Type1 的模型性能提升了 0.47%; Syntax+self-attention-type2 的模型性能比 Self-

attention-Type2 的模型性能提升了 0.19%。

结束语 本文旨在研究隐式篇章关系分类任务,基于 BiLSTM 网络框架,提出了基于自注意力机制和句法信息的神经网络隐式篇章关系分类模型。实验结果表明,本文所构建的系统能够获得比基准系统基于 BiLSTM 网络的隐式篇章关系分类模型更好的效果。

从实验结果可以看出,隐式篇章关系分类任务仍然具有很大的挑战,目前的方法取得的效果还非常有限,所能达到的最好性能还不足 60%。下一步工作中,我们将从两个方面开展研究工作:1)重点研究隐式篇章关系中的具体的语法形式,通过探究数据中的一些隐藏属性来提升模型的性能;2)通过设计新的神经网络模型来提取出更有效的模型特征,让模型获得更好的效果。

参考文献

- [1] POPESCU-BELIS A, MEYER T. Using Sense-Labeled Discourse Connectives for Statistical Machine Translation[C]// Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Pennsylvania, USA: Association for Computational Linguistics, 2012: 129-138.
- [2] JANSEN P, SURDEANU M, CLARK P. Discourse Complements Lexical Semantics for Non-factoid Answer Reranking[C]// Proceedings of the Association for Computational Linguistics. Pennsylvania, USA: Association for Computational Linguistics, 2014: 977-986.
- [3] LOUIS A, JOSHI A, ENKOVA A. Discourse Indicators for Content Selection in Summarization[C]// Proceedings of the Special Interest Group on Discourse and Dialogue. Pennsylvania, USA: Association for Computational Linguistics, 2010: 147-156.
- [4] PITLER E, NENKOVA A. Using Syntax to Disambiguate Explicit Discourse Connectives in Text[C]// Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Pennsylvania, USA: Association for Computational Linguistics, 2009: 13-16.
- [5] PRASAD R, DINESH N, LEE A, et al. The Penn Discourse TreeBank 2.0[C]// Proceedings of the International Conference on Language Resources and Evaluation. Paris, France: European Language Resources Association, 2008: 2961-2968.
- [6] EDDY S. Hidden Markov models[J]. Current Opinion in Structural Biology, 1996, 6(3): 361-365.
- [7] RATNAPARKHI A. A Maximum Entropy Model for Part-of-Speech Tagging[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Pennsylvania, USA: Association for Computational Linguistics, 1996: 133-142.
- [8] COLLINS M. Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms[C]// Proceedings of the Annual Meeting of the Association for Computational Linguistics. Pennsylvania, USA: Association for Computational Linguistics, 2002: 1-8.
- [9] CHANG C C, LIN C J. LIBSVM: A library for support vector machines[M]. ACM, 2011: 1-27
- [10] LAFFERTY J, MCCALLUM A, PEREIRA F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]// Proceedings of the International Conference

- on Machine Learning. Massachusetts, USA: The International Machine Learning Society, 2001:282-289.
- [11] PITLER E, LOUIS A, NENKOVA A. Automatic Sense Prediction for Implicit Discourse Relations in Text[C]// Proceedings of the Association for Computational Linguistics. Pennsylvania, USA: Association for Computational Linguistics, 2009:683-691.
- [12] LIN Z H, KAN M Y, NG H T. Recognizing Implicit Discourse Relations in the Penn Discourse Treebank[C]// Proceedings of Empirical Methods in Natural Language Processing. Pennsylvania, USA: Association for Computational Linguistics, 2009:343-351.
- [13] WANG W T, SU J, TAN C L. Kernel Based Discourse Relation Recognition with Temporal Ordering Information[C]// Proceedings of the Association for Computational Linguistics. Pennsylvania, USA: Association for Computational Linguistics, 2010:710-719.
- [14] RUTHERFORD A, XUE N W. Discovering Implicit Discourse Relations Through Brown Cluster pair Representation and Coreference Patterns[C]// Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. Pennsylvania, USA: Association for Computational Linguistics, 2014:645-654.
- [15] QIN L H, ZHANG Z S, ZHAO H. Shallow Discourse Parsing Using Convolutional Neural Network[C]// Proceedings of the Conference on Computational Natural Language Learning-Shared Task. Pennsylvania, USA: Association for Computational Linguistics, 2016:70-77.
- [16] SCHENK N, CHIARCOS C, DONANDT K, et al. Do We Really Need All Those Rich Linguistic Features? A Neural Network-Based Approach to Implicit Sense Labeling[C]// Proceedings of the Conference on Computational Natural Language Learning-Shared Task. Pennsylvania, USA: Association for Computational Linguistics, 2016:41-49.
- [17] WEISS G, BAJEC M. Discourse Sense Classification from Scratch using Focused RNNs[C]// Proceedings of the Conference on Computational Natural Language Learning-Shared Task. Pennsylvania, USA: Association for Computational Linguistics, 2016:50-54.
- [18] CHEN J F, ZHANG Q, LIU P F, et al. Implicit Discourse Relation Detection via a Deep Architecture with Gated Relevance Network[C]// Proceedings of the Association for Computational Linguistics. Pennsylvania, USA: Association for Computational Linguistics, 2016:1726-1735.
- [19] DOZAT T, MANNING C D. Deep Biaffine Attention for Neural Dependency Parsing[C]// Proceedings of 5th International Conference on Learning Representations. 2017:24-26.
- [20] ZHANG B, SU J, XIONG D, et al. Shallow Convolutional Neural Network for Implicit Discourse Relation Recognition[C]// Proceedings of Empirical Methods in Natural Language Processing. Pennsylvania, USA: Association for Computational Linguistics, 2015:2230-2235.
- [21] RUTHERFORD A, XUE N. Improving the Inference of Implicit Discourse Relations via Classifying Explicit Discourse Connectives[C]// Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Pennsylvania, USA: Association for Computational Linguistics, 2015:799-808.
- [22] LIU Y, LI S. Recognizing Implicit Discourse Relations via Repeated Reading: Neural Networks with Multi-Level Attention[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Pennsylvania, USA: Association for Computational Linguistics, 2016:1224-1233.
- [23] LIU Y, LI S, ZHANG X, et al. Implicit discourse relation classification via multi-task neural networks[C]// Thirtieth AAAI Conference on Artificial Intelligence. USA: AAAI Press, 2016:2750-2756.
- [24] LAN M, WANG J, WU Y, et al. Multi-task Attention-based Neural Networks for Implicit Discourse Relationship Representation and Identification[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Pennsylvania, USA: Association for Computational Linguistics, 2017:1299-1308.