

SaGE: 基于句法感知图卷积神经网络和ELECTRA的中文隐喻识别模型

张声龙¹, 刘颖^{1*}, 马艳军²

1.清华大学人文学院, 中国语言文学系

2.北京百度网讯科技有限公司深度学习技术平台部

zsl18@mails.tsinghua.edu.cn yingliu@mail.tsinghua.edu.cn

mayanjun02@baidu.com

摘要

隐喻是人类语言中经常出现的一种特殊现象, 隐喻识别对于自然语言处理各项任务来说具有十分基础和重要的意义。针对中文领域的隐喻识别任务, 我们提出了一种基于句法感知图卷积神经网络和ELECTRA的隐喻识别模型(Syntax-aware GCN with ELECTRA, SaGE)。该模型从语言学出发, 使用ELECTRA和Transformer编码器抽取句子的语义特征, 将句子按照依存关系组织成一张图并使用图卷积神经网络抽取其句法特征, 在此基础上对两类特征进行融合以进行隐喻识别。我们的模型在CCL 2018中文隐喻识别评测数据集上以85.22%的宏平均F1分数超越了此前的最佳成绩, 验证了融合语义信息和句法信息对于隐喻识别任务具有重要作用。

关键词: 隐喻识别; ELECTRA; 图卷积神经网络; 依存句法

SaGE: Syntax-aware GCN with ELECTRA for Chinese Metaphor Detection

Shenglong Zhang¹, Ying Liu^{1*}, Yanjun Ma²

1.Tsinghua University, Beijing, China

2.Baidu Inc., Beijing, China

zsl18@mails.tsinghua.edu.cn yingliu@mail.tsinghua.edu.cn

mayanjun02@baidu.com

Abstract

Metaphor is a special phenomenon in human languages. It's fundamental and crucial to detect metaphors for many NLP tasks. As for Metaphor Detection in Chinese, we come up with SaGE(Syntax-aware GCN with ELECTRA) which is inspired by linguistics. SaGE utilizes ELECTRA and Transformer encoder to extract the semantic feature of a sentence, and syntactic feature through GCN, which copes with a graph constructed by dependency parsing result. The model concatenates the two features to detect metaphors. SaGE obtains a substantial improvement over the best reported score in CCL 2018 Chinese Metaphor Detection Task Dataset with an 85.22% macro-F1 score. It demonstrates the importance to incorporate both semantics and syntax in Metaphor Detection.

Keywords: Metaphor Detection, ELECTRA, GCN, Dependency Parsing

1 引言

隐喻(Metaphor)是一种比喻性的语言。Lakoff和Johnson (1980)的概念隐喻理论从认知语言学的角度对隐喻现象进行了深入的剖析。隐喻是通过概念映射(Conceptual Mapping)这一机制实现的,在隐喻机制中,一事物(源域)可以用来表示另一事物(目标域),映射的依据则是二者之间的相似性。以句子“书是人类进步的阶梯。”为例,“阶梯”在句子中是源域,“书”则是目标域,二者之间的共同特质是“使人进步”,从而形成了隐喻。隐喻在日常生活中普遍存在,它不仅存在于自然语言之中,更与人的思维和行为密切相关。

识别句子中的隐喻对于自然语言处理来说十分重要,情感分析、机器翻译、问答系统和机器阅读理解等任务都需要模型能处理句子中的字面义和隐喻义,否则可能会影响系统的性能。以英文句子“It seems that he will kick the bucket soon.”为例,若机器不能识别出“kick the bucket”是“去世”的隐喻用法,翻译的结果则会与原句的语境义大相径庭。同时,能够读懂句子中的言外之意,也是机器智能的一种重要体现。

针对中文领域的隐喻识别任务,我们提出了一种基于句法感知图卷积神经网络和ELECTRA的隐喻识别模型(Syntax-aware GCN with ELECTRA, SaGE)。具体而言,模型分为语义处理模块和句法处理模块两个子模块。语义处理模块对ELECTRA语言模型的输出进行处理,抽取出句子的语义信息;句法处理模块则将句子按照依存句法分析的结果组织成一张图,使用图卷积神经网络(Graph Convolutional Network, GCN)抽取句子的句法信息。最终,模型将两个子模块的输出进一步进行融合处理以完成分类。本文的贡献如下:

- 1.针对中文的特点,我们考虑了字和词两种粒度的信息,使用ELECTRA语言模型和GCN分别对句子的语义和句法进行表示,有效增强了模型的性能。我们的模型在CCL 2018中文隐喻识别数据集上超越了此前的最佳成绩,同时较BERT(Bidirectional Encoder Representations from Transformers)、Ernie(Enhanced Representation through Knowledge Integration)和XLNet等基线模型有显著提升。特征消融实验表明,在基线模型的基础上结合SaGE的句法处理模块,能进一步增强语言模型处理隐喻的能力。

- 2.我们是将图卷积神经网络融入中文隐喻识别任务中的首项研究。此前的研究提取出来的各项特征是相对独立的,而SaGE将句子表示为依存图,因此可以将不同的特征联系起来。同时,我们在GCN的基础上进行了拓展,为节点之间的连边赋予了特征。

- 3.从语言学角度为模型进行了解释分析,结合实验结果验证了句法结构对于句子深层语义理解的重要作用。

2 相关研究

如何识别句子中的隐喻一直以来都吸引着诸多学者的研究兴趣。贾玉祥和俞士汶 (2011)利用同义词词林和HowNet作为词典资源计算词语之间的语义关联性,识别中文的名词隐喻句。曾华琳等 (2016)基于最大熵模型提出了一种特征自动选择方法,综合考虑了字符、句法结构和语义信息等不同层次的特征,该方法能在隐喻识别任务中为输入的文本匹配最优的特征模板。张冬瑜等人 (2020)使用BERT模型替代词向量进行语义表示,利用Transformer进行特征提取并通过分类器进行名词隐喻识别。

Mao等 (2019)从语言学理论出发,使用静态的GloVe词向量表示词的本义,使用动态调整的ELMo词向量表示词的语境义,通过神经网络比较二者之间的语义差别进一步进行隐喻识别。Mu等人 (2019)认为理解隐喻要立足于更加具体的语境,他们跳出了短语或句子层级,从更大的范围来处理隐喻,仅使用简单的梯度提升模型便取得了不错的效果。Su等 (2020)将隐喻识别转化为阅读理解问题,他们在BERT原输入的基础上,额外增加了全局语境特征、局部语境特征、问句特征和词性特征等,增强了模型的表示能力,在2020 ACL隐喻识别任务上取得了最佳表现。Gong等 (2020)使用RoBERTa对语言进行向量化表示,捕捉句子的语境信息,同时结合词汇抽象程度、词性标注、主题模型、WordNet 和VerbNet 等外部特征进行判别,进一步提升了模型的性能。

以上诸多研究都考虑了外部特征对于隐喻识别的作用。但是,它们抽取的特征是彼此独立的,没有按照句法规则将各类特征统一起来,忽略了句法结构对于句子的支撑作用。此前已有许多研究者使用非线性结构来处理文本。Tai等人 (2015)将LSTM拓展到了树结构,提出了Tree-LSTM模型。作者利用句法分析的结果将句子组织成句法树,使用Tree-LSTM进行文本情感分析,取得了不错的结果。Marcheggiani和Titov (2017)针对LSTM结构没有利用句法信息

的问题，在双向LSTM层上堆叠了一个GCN层以编码句法信息，他们的模型在语义角色标注任务上表现出色。Tao Ji等人 (2019)使用图神经网络进行依存句法分析，使得依存树节点能够有效地表示高阶信息，其方法在多个公开数据集上表现优异。受到此前研究的启发，本文在提取特征时将一个句子中的词按照依存关系连接起来，组成一张图，在此基础上使用图卷积神经网络对句子进行处理。

3 SaGE模型

当前主流的中文预训练语言模型(如BERT等)大多是基于字粒度的，它们在诸多自然语言处理任务中表现优异，展现了强大的语义编码能力，这对于隐喻识别任务而言是不可或缺的。另一方面，隐喻识别需要考虑某些外部句法特征，而这些句法特征往往是基于词粒度的。为了将字粒度的语义信息和词粒度的外部句法特征结合起来，增强模型的表示能力，我们设计了一种基于句法感知图卷积神经网络和ELECTRA的隐喻识别模型(SaGE)。

SaGE模型结构如图1所示。该模型可以分为两大模块，左侧为语义处理模块，右侧为句法处理模块。语义处理模块中，句子 s 被切分为 t_1, t_2, \dots, t_n 共 n 个字符，它们经过ELECTRA语言模型处理后，输出为字向量 e_1, e_2, \dots, e_n 。字向量再经由Transformer编码器进行进一步处理以抽取全局语境信息，输出句子的语义特征。句法处理模块中，句子 s 被分词器切分为 w_1, w_2, \dots, w_m 共 m 个词，并将每个词的词性嵌入为低维向量，记为 p_{w_i} 。同时，句子可以被依存句法分析器解析成一张依存图，若两个词之间存在依存关系，则为它们建立一条边，并将每一条边上的依存关系嵌入为低维向量，记为 d_{e_j} 。图卷积神经网络会对句子的依存图进行处理，输出图的嵌入向量作为句法特征。进一步，模型将语义特征向量和句法特征向量拼接起来并输入到分类器中，完成分类。

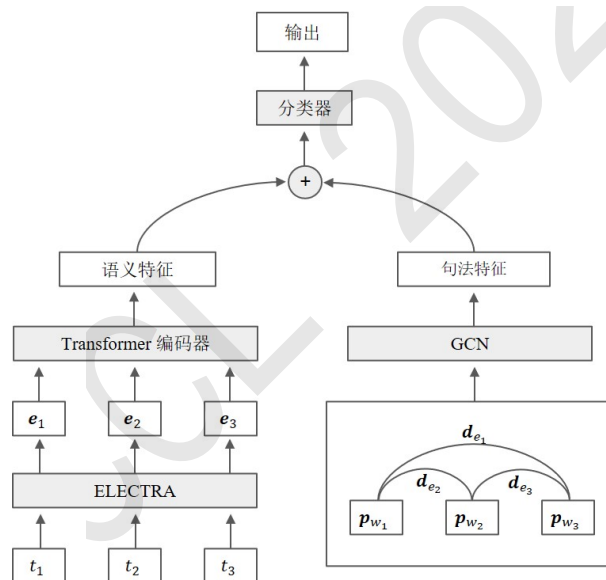


图 1: SaGE模型结构

3.1 ELECTRA语言模型

ELECTRA是Google和斯坦福大学 (Kevin Clark et al., 2020)于2020年提出的一种全新的预训练语言模型，它采用Transformer编码器作为基础组件。不同于BERT等掩码语言模型(Masked Language Model)，ELECTRA提出了一个新颖的训练任务。

在掩码语言模型中，以BERT为例，模型会随机遮盖掉一部分词，并通过后续的训练试图还原被遮盖的词，这种训练方式存在着一定的缺陷：首先，掩码标记“[mask]”是人工标注的，这在后续的下游任务中实际上并不会出现，因此可能会造成一定的性能损失；其次，BERT只对句子中15%的词进行随机处理，模型训练过程中只对其中15%的词进行了预测，没有充分利用大规模语料的优势。

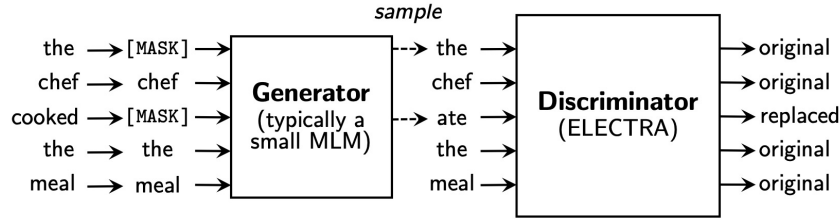


图 2: ELECTRA训练思路

ELECTRA借鉴了对抗生成网络(Generative Adversarial Networks, GAN)的思想, 引入了判别式任务(Discriminative Task)。第一步, 模型通过生成器(Generator, 通常是一个小型的掩码语言模型)与BERT一样对被破坏的句子进行还原, 即预测被随机遮盖的词。第二步, 模型通过判别器(Discriminator)对生成器输出的句子进行处理, 判定其中的词是原来的还是被生成器替换的。显然, ELECTRA在训练过程中预测了句子中的全部词, 克服了掩码语言模型只预测部分词的缺点。在训练完成后, 使用判别器处理下游任务, 因此, 解决了实际任务中“[mask]”不出现的问题。实验证明, ELECTRA在同等模型体积、数据量和算力的情况下比BERT表现要更为优秀。

在我们的隐喻识别模型中, ELECTRA会对每一个输入的句子 S 进行处理, 为句中每一个字生成 d 维的字向量。设 S 共包含 n 个词, 则句子可用向量 $S = (t_1, t_2, \dots, t_n)$ 表示。其中, t_i 为句中第 i 个字符。ELECTRA的运算可以简单表示为公式(1)。

$$E_s = ELECTRA(S), \quad E_s \in \mathcal{R}^{n \times d} \quad (1)$$

E_s 为整个句子的嵌入矩阵, 矩阵的每一行为句中每一个字符的字向量。利用ELECTRA强大的语义建模能力, 我们能很好地表示句中的词在当前语境下的语义, 从而在语义层面上克服隐喻识别的障碍。

3.2 Transformer编码器

Transformer编码器来自于Google (Ashish Vaswani et al., 2017)发布于2017年的一项工作, 它创造性地抛弃了此前在自然语言处理任务中广泛使用的循环神经网络(Recurrent Neural Network, RNN)或卷积神经网络结构(Convolutional Neural Network, CNN), 代之以自注意力(Self-Attention)机制, 事实证明, Transformer相较于RNN和CNN等结构有着明显的优势。一方面, 由于注意力机制的使用, 模型可以关注更大范围内的语境信息, 避免句子过长导致的网络退化。另一方面, Transformer可以很方便地利用现有的主流设备高效并行地工作, 大大提升了运算效率。如今, Transformer已经成为BERT、XLNet和Ernie等大型预训练语言模型采用的基础组件。

如图3所示, 在我们的模型中, Transformer编码器的输入由两部分组成, 一是来自于上游ELECTRA语言模型产生的字向量, 它们携带着丰富的语义信息。二是句子中每一个字对应的位置编码, 因为在自然语言处理任务中, 考虑字符的先后顺序是十分重要的。字向量和位置编码相加, 共同组成编码器的完整输入。在多头注意力(Multi-Head Attention)环节, 编码器会考虑每个字在句子中的上下文, 计算句子中当前字对于包括其自身在内的所有字符的注意力, 从而实现对整个句子的语义信息抽取, 更好地建模句子的语义信息。此外, 编码器中还有残差连接和归一化层(Add&Norm), 二者的目的都是防止网络退化。

在实验中, 我们利用Transformer编码器对ELECTRA输出的句嵌入矩阵 E_s 进行处理。其运算过程如下:

$$H = TE(E_s), \quad H \in \mathcal{R}^{n \times k} \quad (2)$$

$$x^h = \text{Mean}(H), \quad x^h \in \mathcal{R}^k \quad (3)$$

$$x^{sem} = x^h W + b, \quad x^{sem} \in \mathcal{R}^\alpha \quad (4)$$

公式(2)的 H 为句嵌入矩阵 E_s 经过Transformer编码器后输出的隐藏状态矩阵, 矩阵的每一行表示句中每一个词的隐藏状态。公式(3)中, 我们对 H 取平均, 压缩为 k 维向量 x^h , 其

中 $Mean$ 表示取平均操作。 x^h 再经(4)中的线性变换输出为 α 维向量，得到 x^{sem} ，即为语义特征向量。Transformer编码器使得模型能够将关注范围扩大至整个句子，实现句层级的语义信息提取，有助于发掘源域和目标域之间的语义差异。

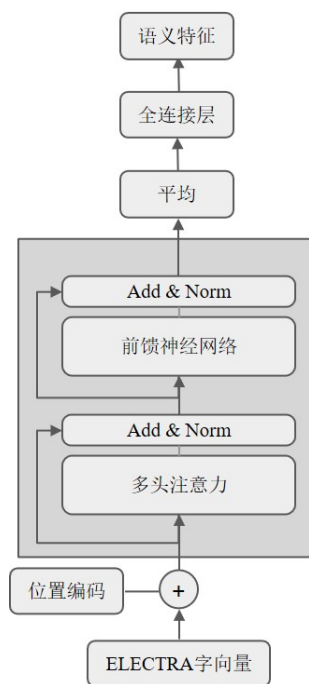


图 3: Transformer编码器

3.3 图卷积神经网络

传统的机器学习方法在欧氏数据上取得了重要的突破。文本和图片都是属于欧氏空间的数据，其特点就是结构十分规则。而在现实生活中，有许多不规则的数据结构，其中以图结构(Graph)为典型。图中每个节点周围的结构可能都是独一无二的。社交网络、化学分子结构和知识图谱等，都可以表示为图结构。

以树结构或图结构表示句子具有一定的合理性。一方面，句子内部是有层次的，人类很容易察觉到句子的层次性，然而机器却只能在单一维度上向前或向后看，不符合人对语言的理解方式。另一方面，图神经网络可以灵活编码不同的语言特征。若图中的每一个节点表示一个词，每一条边表示词与词之间的关系，则节点可以嵌入如词义、词性和实体信息等不同方面的知识；图中的边则可以嵌入依存关系(Dependency)、句法成分结构(Constituent)等不同的关系，对语言进行更加准确和丰富的特征表示。自然语言中的句子可以根据句法规则组织成树形结构，而树结构本质上是一种特殊的图结构，因此，将句子组织成一张图，是一种自然的想法。

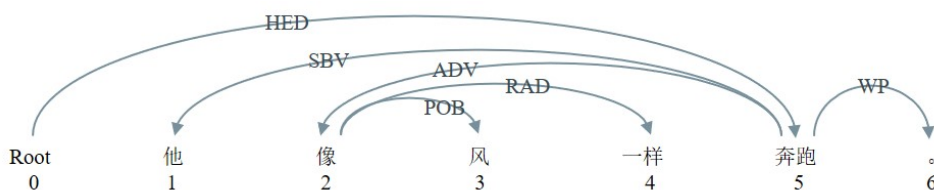


图 4: 依存图

以句子“他像风一样奔跑。”为例，该句子可以被句法分析器解析为如图4所示的依存关系图，每个词都是依存图中的一个节点，其中“Root”为虚拟的头节点，从“Root”开始逐次给每个词编号。我们按照依存边的指向为节点添加边，最终构成一张完整的图。此外，我们将词的词

性标注嵌入为低维向量，作为节点特征；若节点与节点之间存在一条边，则将边上的依存关系也嵌入为低维向量，作为边特征。

GCN将卷积神经网络的思想迁移到了图结构上。Gilmer等人 (2017)提出了一种名为消息传递图神经网络(Message Passing Neural Network, MPNN)的框架，在MPNN框架下，节点可以发出消息，也可以通过节点之间的连边接收其邻居发出的消息。我们基于MPNN构建空域图卷积神经网络，以便对有向图进行处理。

对于图 $G = (V, E)$ ， V 表示节点的集合， E 为边的集合。对每一个节点 i ，记其邻居为 j ， $j \in N_i$ ，节点 i 与邻居 j 之间的边为 e_{ij} 。节点 i 有特征 x_i ，边 e_{ij} 上有特征 y_{ij} 。第 $l + 1$ 次卷积运算中，节点 i 的隐藏状态 x_i^{l+1} 的更新逻辑可用公式(5)表示：

$$x_i^{l+1} = \sigma \left(\sum_{j \in N_i} \left([x_j^l; y_{ij}^l] W^l + b^l \right) \right) \quad (5)$$

在第 $l + 1$ 次卷积操作时，我们将当前邻居节点的特征 x_j^l 和二者的边特征 y_{ij}^l 拼接起来($[x_j^l; y_{ij}^l]$ 表示拼接操作)，乘以权重 W^l ，并加上此处的偏置 b^l 。最后，我们再对结果施以非线性函数 σ (在GCN中通常使用ReLU)，用得到的结果更新节点 x_i 的状态，记为 x_i^{l+1} 。

简单来说，在每一次卷积操作时，节点 i 都会对其所有的邻居传递过来的消息进行处理，并以此为依据更新自己的值。第一次卷积时，节点可以接收到邻居的消息，第二次卷积时则可以接收到邻居的邻居的消息，因为此时的邻居在上一次卷积时同样也接收到了来自它的邻居的消息。从而，经过多次卷积之后，节点可以考虑到与自己距离较远的信息。

回到句子层级。设输入GCN的句子为 S ，句中一共有 m 个词，GCN第 $l + 1$ 层输出的词的隐藏状态为 β 维。则 S 经过第 $l + 1$ 层后可用公式(6)表示：

$$H^{l+1} = \begin{pmatrix} x_1^{l+1} & x_2^{l+1} & \dots & x_m^{l+1} \end{pmatrix}, \quad H^{l+1} \in \mathcal{R}^{m \times \beta} \quad (6)$$

其中， x_i^{l+1} 为句中第 i 个词在第 $l + 1$ 层时的隐藏状态，其计算方式依公式(5)。最后，我们对所有节点的隐藏状态取平均，输出为向量 x^{syn} ，作为整个依存图的嵌入表示，此即模型所抽取的句法特征向量。如公式(7)：

$$x^{syn} = Mean(H^{l+1}), \quad x^{syn} \in \mathcal{R}^{\beta} \quad (7)$$

3.4 分类器

深度学习中通常使用全连接层作为分类器。我们将模型上游传递过来的语义特征 x^{sem} 和句法特征 x^{syn} 进行拼接，作为这个句子最终的特征化表示，即句特征，记为 x ， $x \in \mathcal{R}^{\alpha+\beta}$ 。该层的计算逻辑可用公式(8)表示。

$$\begin{aligned} y &= xW + b \\ &= [x^{sem}; x^{syn}]W + b \end{aligned} \quad (8)$$

其中， W 为此处的权重， b 为偏置。句特征经过全连接层后输出为向量 y ， $y \in \mathcal{R}^c$ ，其中 c 为分类任务的类别数。

4 实验

4.1 数据集及评价指标

本实验采用大连理工大学于CCL 2018中文隐喻识别与情感分析任务中发布的中文隐喻识别数据集¹。该数据集分为训练集与测试集两部分，其概况如表1所示。

¹<https://github.com/DUTIR-Emotion-Group/CCL2018-Chinese-Metaphor-Analysis>

	样本数	总字数	总词数	平均句长	最大句长	最小句长
训练集	4394	81450	53564	18.54	246	4
测试集	1100	20970	13746	19.07	144	4

注：其中句长均以字符为单位进行统计

表 1: 数据集概况

数据集一共包含三类样本，即动词隐喻句、名词隐喻句和非隐喻句。我们可以从语义冲突的观点来理解不同类型的隐喻，它是隐喻产生的基本条件。语义冲突也可称为语义偏离，指在语言意义组合中违反语义选择限制或常理的现象 (束定芳, 2002)。

我们根据句子的句法特征，以表2为例对动词隐喻句、名词隐喻句和非隐喻句三类句子进行详细说明 (束定芳, 2000)。

类别	举例
名词隐喻句	1. 被窝像暖炉一样，热烘烘的。 2. 波光粼粼的水面是灿烂的银河。
动词隐喻句	3. 朋友的鼓励点燃了他的热情。 4. 台风重创了这个多灾多难的城市。
非隐喻句	5. 杰克对中华文化十分感兴趣。 6. 深度学习的发展令人瞩目。

表 2: 隐喻句分类

名词隐喻指的是由名词构成的隐喻。句1中的源域是“暖炉”，目标域是“被窝”；句2中的源域是“银河”，目标域是“水面”。两个句子中作为源域和目标域的两个名词属于不同范畴，描述的是不同的实体，因此产生了语义上的冲突，形成了名词隐喻句。

动词隐喻指的是话语中使用的动词于逻辑上的主语或宾语构成语义冲突所形成的隐喻。句3中动词“点燃”一般用于可燃物，而非抽象的概念“热情”；句4中动词“重创”多用于人或动物，此处用于无生命的“城市”。两个句子中的动词促使听者从冲突的语义出发去感知其引申义，由此构成了动词隐喻句。

所谓非隐喻句，指的是句中的词语没有产生语义冲突，未触发隐喻机制。句5和句6中皆为非隐喻句。

数据集中三类隐喻句的分布情况如表3所示。

	名词隐喻句数	动词隐喻句数	非隐喻句数
训练集	2035	2040	319
测试集	507	517	76

表 3: 样本分布

我们的实验本质上是一个针对句子的三分类任务。依照CCL 2018的评测标准，我们同样采用宏平均F1值(macro-F1)作为实验的评价标准。宏平均首先对每一个类*i*的预测结果计算精确率(P)、召回率(R)以及F值，然后对所有类的F值求算术平均值。具体公式如下：

$$F_i = \frac{2 \times P_i \times R_i}{P_i + R_i} \times 100\% \quad (9)$$

$$Macro_F = \frac{1}{n} \sum_{i=1}^n F_i \quad (10)$$

4.2 基线模型

在实验中，我们选取了经典的语言模型BERT (Devlin Jacob et al. , 2018)、Ernie (Sun Yu et al. , 2019)和XLNet (Zhilin Yang et al. , 2019)作为基线模型(Baseline)以验证本文方法的有效性。其中，BERT采用Google发布的原版bert-base-chinese，Ernie使用百度公司发布的Ernie 1.0，XLNet和本文模型中使用的ELECTRA采用哈工大发布的开源预训练语言模型XLNet-base和ELECTRA-discriminator-base (Yiming Cui et al. , 2020)。

此外，我们也选择了在CCL 2018中文隐喻识别评测任务中排名前三名的队伍(分别记为CCL 2018 top 1, CCL 2018 top 2 和CCL 2018 top 3)作为参照(赵晓妮, 2019)。由于评测任务中各个队伍使用的方法已不可考，我们无法进行模型上的详细说明，故仅从实验结果上验证本文方法的有效性。

4.3 实验细节

数据预处理：使用哈工大发布的LTP 4.0工具 (Wanxiang Che et al. , 2020)对数据集中的全部句子进行词性标注和依存句法分析。进一步，将每个词的词性嵌入为8维向量，词与词之间的依存关系也嵌入为8维向量。

ELECTRA：使用ELECTRA模型的判别器作为语言模型对句子进行向量化表示，句中每一个字均表示为768维的向量。

Transformer编码器：仅使用一个编码器层和一个注意力头。编码器对上游传递过来的ELECTRA字向量进行处理，对句子的全部字向量取平均，然后通过一个全连接层输出为64维的向量，以此提取句子的语义信息，记为语义特征向量。

图卷积神经网络：使用DGL图神经网络框架 (Mingjie Wang et al. , 2020)，按照依存关系构建有向图。实验中GCN包含两个卷积层，第一个卷积层的输出维度为8；第二个卷积层的输出即句子的句法特征向量，其维度会对结果产生一定的影响。我们进行了多轮实验，取表现最优的情况：在基于ELECTRA和BERT的实验中，我们取输出维度为3，在基于XLNet和Ernie的实验中则取为4。

分类器：使用全连接层，其输入为语义特征向量和句法特征向量拼接后的句特征向量；输出维度为3，即分类的总类别数。

参数设置：使用dropout策略，其p值设置为0.3。使用带Softmax的交叉熵损失函数。优化器使用AdamW，初始学习率设置为 $5e-5$ 。Batch Size设置为16。

实验环境：编程语言采用Python 3.8，深度学习框架为PyTorch 1.7。CPU为Intel 10700K，使用一块NVIDIA RTX 3080 GPU。

4.4 实验结果

我们的实验分为两部分，分别是与基线模型的对比及特征消融实验。

基线模型对比：对于BERT、Ernie和XLNet，我们仿照SaGE的框架，但去掉了模型右侧的句法处理模块，即只在各个语言模型的基础上堆叠一个单层单注意力头的Transformer编码器，并输出为64维的向量，再通过一个全连接层完成分类。

特征消融：目的是验证句法处理模块的有效性。一方面，我们在前面构建的基线模型的基础上增加了SaGE模型右侧的句法处理模块，分别记为BERT+*syn*、Ernie+*syn*和XLNet+*syn*。另一方面，去除SaGE的句法处理模块，此时模型退化为原始的ELECTRA，也可记为SaGE-*syn*。

表4所示为所有模型的宏平均F1分数(使用百分制)。我们提出的SaGE模型在和全部基线模型的对比中表现出色，以85.22%的宏平均F1分数超越了CCL 2018中文隐喻识别评测任务中的第一名约2个百分点。此外，此时的准确率(Accuracy)可以达到89.45%。同时可以看到，单纯地使用XLNet、BERT和Ernie效果并不理想，甚至无法超越此前的最佳模型。这说明在隐喻识别中充分考虑到句子的内部结构、利用外部资源等具有一定的现实意义，而这也是我们研究的出发点所在。

如表5所示。XLNet、BERT和Ernie等基线模型在结合了本文提出的句法处理模块后，性能都有了明显的提升，宏平均F1值分别上升了1.91、1.57和1.24个百分点。此外，当去除SaGE的句法处理模块后，模型的宏平均F1值下降了0.94个百分点。特征消融实验验证了句法信息对于隐喻识别的重要性，我们设计的句法处理模块是有效的。

模型	macro-F1(%)
XLNet	80.73
BERT	81.57
Ernie	82.10
CCL 2018 top 3	82.72
CCL 2018 top 2	83.09
CCL 2018 top 1	83.27
SaGE	85.22

表 4: 基线模型对比

模型	macro-F1(%)
XLNet+ <i>syn</i>	82.64 (+1.91)
BERT+ <i>syn</i>	83.14 (+1.57)
Ernie+ <i>syn</i>	83.34 (+1.24)
ELECTRA (SaGE- <i>syn</i>)	84.28 (-0.94)

表 5: 特征消融

虽然预训练语言模型改变了自然语言处理的研究范式，在许多任务中用户仅需要调用接口并微调就可以取得不错的结果，但是语言本身是复杂的，它包含着语义、语音、句法和修辞等多个方面，近年来兴起的多模态自然语言处理也正是试图从多个层面对语言进行建模。

事实上，已经有许多研究说明了句法结构对于隐喻识别的重要性。Sullivan (2013)从构式语法的角度出发，发现在许多情况下，特定的构式决定了何种句法成分能够充当隐喻中的源域和目标域。其中，动词往往会激活句子中的源域，目标域则由动词的论元所激活。Hwang (2014)对语言中的使役结构(Caused Motion)进行了研究，发现动词的补足语往往呈现为隐喻表达。

因此，我们选择使用依存句法分析有以下考虑：一是依存句法分析(Dependency Parsing)相较于句法成分分析(Constituent Parsing)准确率更高，能最大程度减少上游预处理带来的错误累积；二是依存句法分析中，动词居于句子的主导地位，既能以一种相对合理的方式将句子组织成图结构，又能充分地考虑动词对句子隐喻义表达的影响。

4.5 偏误分析

为了进一步观察模型的性能，明确SaGE在处理隐喻识别任务时的缺陷，我们提取出了测试集中模型分类错误的样本，并进行了仔细的研究。实验中的偏误大致可以分成三类，具体例子如表6所示。

类别	例句	标注标签	预测标签
兼类	1. 却不能阻止你像不倒松柏一样一直挺立。	动词隐喻	名词隐喻
	2. 眼睛常常说出心的话语。	名词隐喻	动词隐喻
标注错误	3. 夕阳西下，天空宛若一幅油画。	动词隐喻	名词隐喻
	4. 淅淅沥沥的一场春雨，滋润着牧奎村的黑土地，也淋湿了老张的心。	非隐喻	动词隐喻
预测错误	5. 经过仔细排查，犯罪分子终于浮出水面。	动词隐喻	非隐喻
	6. 几星萤火游来游去，不像飞行，像在厚密的空气里漂浮。	动词隐喻	名词隐喻

标注标签：数据集中标注者人工给定的标签
 预测标签：SaGE通过运算预测的标签

表 6: 偏误举例

兼类：指句子既有名词隐喻的性质，又有动词隐喻的特点，在判断时是模棱两可的。如句1，将“你”比作“松柏”可以视为名词隐喻，但用来描述人的动词“挺立”用于“松柏”上，又构成了动词隐喻。再如句2，“眼睛”无法发生“说”的动作，在这一层面上该句属于动词隐喻句；然而，“心”又无法产生“话语”，“心的话语”实则指“心声、想法”，此处又形成了名词隐喻。这类句子难以用某一特定的类别进行分类，因此，在给隐喻句进行分类时，我们应该制定更加详细的标准，以处理兼属两种隐喻句的句子。

标注错误: 由于隐喻识别数据集的构建很大程度上依赖于人工标注, 因此可能会出现少量难以避免的标注错误。如句3, 将“天空”比作“油画”是很显然的名词隐喻句, 前者是目标域, 后者是源域。再如句4, “淋湿”和“心”之间有十分明显的语义冲突, 应该属于动词隐喻句。我们的模型实际上在这两个例子中给出了正确的答案。因此, 在数据标注上, 我们应进一步明确标注规范, 尤其是要加强后期的人工审校。

预测错误: 模型本身处理不当而造成的错误。该类错误主要包括两种。其一是某些需要背景知识的语言表达, 如句5, “浮出水面”指“身份揭晓”, 此处模型由于缺少相关的知识, 很容易将其理解为字面意义。第二种是以比喻词“像”、“似的”和“犹如”等引导的隐喻句, 由于该类比喻词较多出现于名词隐喻句中, 因此在数据量上可能会给模型带来一定的诱导, 当出现如句6一样较为复杂的情况时, 模型会倾向于将其理解为名词隐喻。针对模型缺乏背景知识的问题, 我们可以进一步在模型中融入外部知识, 如词汇抽象程度、HowNet和同义词词林等。针对模型对比喻词引导的句子处理较差的问题, 一方面要考虑样本类别之间的平衡性, 尽可能做到每一种隐喻句的数量大体一致, 对于样本稀少的类别, 可以进行数据增强; 另一方面, 可以将规则和深度学习的方法结合起来, 用人工编写的规则来处理包括该类隐喻句在内的较为复杂的情况。

5 总结与展望

本文针对中文设计了一种基于句法感知图卷积神经网络和ELECTRA的隐喻识别模型, 在CCL 2018中文隐喻识别任务数据集上取得了不错的效果, 验证了将句法结构融入到隐喻识别任务中的有效性, 并结合语言学理论给出了相应的解释。此外, 我们对模型预测错误的样本进行了深入分析, 对隐喻识别任务中存在的几类问题进行了探讨, 为今后的研究者开展相关的探索提供了新的思路。

由于时间精力有限, 我们的研究尚存在诸多的不足。首先, 我们从利用非线性结构进行自然语言处理的研究中得到启发, 使用GCN对依存句法分析的结果进行处理, 但采用的网络结构较为单一。未来, 我们将探索其他的图神经网络结构, 如图注意力网络(Graph Attention Network, GAT)和关系图卷积神经网络(Relational Graph Convolutional Network, R-GCN)等。其次, 我们将进一步探索隐喻识别数据集的数据增强方法, 扩大数据集的规模; 在英文数据集中拓展我们的研究, 探讨中英文隐喻识别的异同。最后, 尝试将HowNet和同义词词林等外部知识融入到中文隐喻识别任务中来, 发挥语言学知识在自然语言处理中的积极作用。

致谢

本文工作由清华大学人文社科振兴项目“基于大数据技术的明清小说分析和研究”(2019THZWJC38), 教育部人文社会科学研究一般项目“语体特征的自动提取和研究”(17YJAZH056), 国家社会科学基金重大项目“基于大数据技术的古代文学经典文本分析与研究”(18ZDA238)和百度网讯科技有限公司项目“基于深度学习框架PaddlePaddle的开源教程和案例建设”资助。

参考文献

- 贾玉祥, 俞士汶. 2011. 基于词典的名词性隐喻识别. 中文信息学报. 25(02):99-104.
- 束定芳. 2000. 论隐喻的基本类型及句法和语义特征. 外国语(上海外国语大学学报). 2000(01):20-28.
- 束定芳. 2002. 论隐喻的运作机制. 外语教学与研究. 2002(02):98-106+160.
- 曾华琳, 周昌乐, 陈毅东, 史晓东. 2016. 基于特征自动选择方法的汉语隐喻计算. 厦门大学学报(自然科学版). 55(03):406-412.
- 张冬瑜, 崔紫娟, 李映, 张伟, 林鸿飞. 2020. 基于Transformer和BERT的名词隐喻识别. 数据分析与知识发现. 4(04):100-108.
- 赵晓妮. 2019. 基于比喻性修辞的文本情感分析研究. 哈尔滨工业大学.
- Wanxiang Che, Yunlong Feng, Libo Qin and Ting Liu. 2020. N-LTP: A Open-source Neural Chinese Language Technology Platform with Pretrained Models. *arXiv preprint arXiv:2009.11616*.

- Kevin Clark, Minh-Thang Luong, Quoc V. Le and Christopher D. Manning. 2020. ELECTRA: Pretraining Text Encoders as Discriminators rather than Generators. In *ICLR 2020* .
- Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang and Guoping Hu. 2020. Revisiting Pre-Trained Models for Chinese Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 657-668.
- Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals and George E. Dahl. 2017. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1263-1272.
- Hongyu Gong, Kshitij Gupta, Akriti Jain and Suma Bhat 2020. IlliniMet: Illinois System for Metaphor Detection with Contextual and Linguistic Information. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 146-153.
- Jena Hwang. 2014. *Identification and Representation of Caused-Motion Constructions*. University of Colorado.
- Devlin Jacob, Chang Ming-Wei, Kenton Lee and Toutanova Kristina. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .
- Tao Ji, Yuanbin Wu, and Man Lan. 2019. Graph-based Dependency Parsing with Graph Neural Networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2475-2485.
- George Lakoff and Mark Johnson. 1980. *Metaphors we live by*. University of Chicago Press, Chicago, USA.
- Rui Mao, Chenghua Lin, and Frank Guerin. 2019. End-to-End Sequential Metaphor Identification Inspired by Linguistic Theories. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3888-3898.
- Diego Marcheggiani, Ivan Titov. 2017. Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506-1515.
- Jesse Mu, Helen Yannakoudakis and Ekaterina Shutova. 2019. Learning Outside the Box: Discourse-level Features Improve Metaphor Identification. In *Proceedings of NAACL-HLT 2019*, pages 596-601.
- Chuangdong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiyi Li, Rongbo Wang, and Zhiqun Chen. 2020. DeepMet: A Reading Comprehension Paradigm for Token-level Metaphor Detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 30-39.
- Karren Sullivan. 2013. *Frame and Constructions in Metaphoric Language*. Amsterdam: John Benjamins Publishing Company.
- Kai Sheng Tai, Richard Socher, Christopher D. Manning. 2015. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1556-1566.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* .
- Minjie Wang, Da Zheng, Zihao Ye, Quan Gan, Mufei Li, Xiang Song, Jinjing Zhou, Chao Ma, Lingfan Yu, Yu Gai, Tianjun Xiao, Tong He, George Karypis, Jinyang Li and Zheng Zhang. 2020. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *arXiv preprint arXiv:1909.01315*.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. 2019. XLNet: Generalized autoregressive pretraining for language understanding. In *NeurIPS* .
- Sun Yu, Wang Shuohuan, Li Yukun, Feng Shikun, Chen Xuyi, Zhang Han, Tian Xin, Zhu Danxiang, Tian Hao and Wu Hua. 2019. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1907.12412*.