

基于GCN和门机制的汉语框架排歧方法*

游亚男^{1,‡}, 李茹^{1,2,*†}, 苏雪峰^{1,3,‡}, 闫智超^{1,‡}, 孙民帅^{1,‡}, 王超^{1,‡}

¹山西大学 计算机与信息技术学院, 山西 太原 030006

²山西大学 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006

³山西工程科技职业大学现代物流学院, 山西 晋中 030609

[‡]{2280493770, 455375251, 751824801, 1070913573, 1055342647}@qq.com

*{liru}@sxu.edu.cn

摘要

汉语框架排歧旨在候选框架中给句子中的目标词选择一个符合其语义场景的框架。目前研究方法存在隐层向量的计算与目标词无关, 并且忽略了句法结构信息对框架排歧的影响等缺陷。针对上述问题, 使用GCN对句法结构信息进行建模; 引入门机制过滤隐层向量中与目标词无关的噪声信息; 并在此基础上, 提出一种约束机制来约束模型的学习, 改进向量表示。该模型在CFN、FN1.5和FN1.7数据集上优于当前最好模型, 证明了方法的有效性。

关键词: 汉语框架排歧; 句法信息; GCN; 门机制

Chinese Frame Disambiguation Method Based on GCN and Gate Mechanism

Yanan You^{1,‡}, Ru Li^{1,2,*†}, Xuefeng Su^{1,3,‡}, Zhichao Yan^{1,‡}, Minshuai Sun^{1,‡}, Chao Wang^{1,‡}

¹School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China

²Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, Shanxi 030006, China

³School of Modern Logistics, Shanxi Vocational University of Engineering Science and Technology, Jinzhong, Shanxi 030609, China

[‡]{2280493770, 455375251, 751824801, 1070913573, 1055342647}@qq.com

*{liru}@sxu.edu.cn

Abstract

Chinese frame disambiguation aims to select a frame that matches its semantic scene for the target word in the sentence among the candidate frames. The current research methods have the defects that the calculation of the hidden layer vector has nothing to do with the target word, and ignores the influence of the syntactic structure information on the frame disambiguation. Aiming at the above problems, GCN is used to model the syntactic structure information; a gate mechanism is introduced to filter the noise information irrelevant to the target word in the hidden layer vector; and on this basis, a constraint mechanism is proposed to constrain the learning of the model and improve the representation vector. The model outperforms the current state-of-the-art models on the CFN, FN1.5 and FN1.7 datasets, proving the effectiveness of the method.

Keywords: Chinese frame disambiguation, Syntactic information, GCN, Gate mechanism

1 引言

* 基金项目: 基于语言认知机理的汉语框架语义计算研究 (61936012)

† 通讯作者 Corresponding Author

©2022 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

汉语框架网 (Chinese FrameNet, CFN) (You and Liu, 2005)是以Fillmore的框架语义学 (Fillmore et al., 1976)为理论基础, 以汉语真实语料为依据, 参照伯克利大学的框架语义知识库 (FrameNet, FN) (Baker et al., 1976)构建的汉语词汇语义知识库, 包括框架库、句子库和词元库, 其中相关术语如表1 (Li, 2012)所示。汉语框架语义分析是基于汉语框架网的标注资源提出的任务 (Shi et al., 2014), 而汉语框架排歧作为汉语框架语义分析中的重要研究任务, 其正确与否直接关系到汉语框架语义分析的准确性。

汉语框架排歧任务即给定一个句子和一个目标词以及该目标词在CFN中能激起的所有框架, 需根据目标词的上下文推断出该目标词在当前语境下所匹配的框架。如表2所示, 在CFN中, 目标词“炒”可以激起两个框架“解雇”、“烹饪”, “解雇”表示雇主终结与雇员之间的雇佣关系; “烹饪”表示烹调者对食物进行加工。在s1中, 根据上下文信息可以判断“炒”激起框架为“解雇”, 而s2中“炒”激起框架为“烹饪”。

术语	定义
目标词 框架	在一个具体的句子中能够激起框架的词元与一些激活性语境相一致的结构化范畴系统, 是存储在人类认知经验中的图示化情境
框架元素	框架的语义参与者, 也称为框架语义角色

表 1: 框架语义分析术语表

词元	候选框架	例句
炒	烹饪、解雇	s1: 如果再这么下去, 老板<炒tgt “解雇”>你没商量。
		s2: 舅妈特地<炒tgt “烹饪”>了过油肉, 迎接我们的到来。

表 2: 语料示例

目前大多数研究将框架排歧看作多分类任务, 主要有两类方法: 一种是基于传统机器学习的方法, 该类方法使用自然语言处理工具分析句子, 人工抽取特征, 通过机器学习方法训练模型。该类方法得到的特征矩阵维度大, 特征稀疏, 并且特征之间没有关联, 模型难以在不同的树结构上并行计算, 计算效率低下。二是基于深度学习的方法, 该类方法实现了自动学习特征, 避免了特征维度过高, 有效的融合了上下文信息, 一定程度上解决了特征之间无关联的问题, 并且已被用于产生该任务的最先进性能。已有工作将依存信息集成到深度学习的模型中用来进一步提高模型性能, 取得了一定的效果。但这些方法仍然存在以下不足: a)在框架排歧中不同目标词在上下文中关注的词应该有所不同, 但现有方法在进行向量计算时并没有考虑这一点。如表2中, 句子s2中, “炒”和“迎接”是两个目标词, 在对它们进行框架排歧时希望模型得到的句子表示是不同的; b)已有方法只考虑句法上与目标词相邻的词, 忽略了句法不相邻词对目标词的影响。

本文首先使用BERT (Devlin et al., 2018)进行上下文编码, 利用哈工大语言技术分析平台 (LTP) (Che et al., 2010)进行句法分析以此构建依存图。然后通过GCN对依存图中的信息交互进行建模, 充分利用GCN对图的特征提取能力。针对a)问题, 考虑到BERT得到的隐层向量并不是特定于目标词的, 输入GCN更新节点的过程中会保留与目标词无关的冗余信息, 因此本文引入门机制, 计算一个特定于目标词的向量, 并将门向量应用于每个GCN层, 改变上下文的表示, 得到特定于目标词的隐层向量来过滤无关特征; 针对b)问题, 本文认为句法不相邻词对于目标词的表示学习具有指导作用, 因此本文提出依据句法信息为句子中的每个词分配一个分数, 明确量化其对目标词进行框架排歧的贡献, 注入模型, 以此来约束模型的学习, 改进表示向量。

本文的贡献之处包括: 1) 提出一种基于GCN和门机制的向量调整方法, 生成特定于目标词的向量表示, 过滤与目标词无关的噪声信息。2) 引入一种基于依存图的约束机制来获取句

子中每个词相对于目标词的重要性得分，注入模型，作为计算隐层向量的训练信号，改进向量表示。3) 在数据集CFN、FrameNet1.5和FrameNet1.7上进行了详细的对比实验，实验结果表明，本文方法有效提高了框架排歧的准确率。

2 相关工作

著名语言学家Fillmore基于认知角度提出了框架语义学。此后，SemEval2007语义评测任务 (Baker et al., 2007) 提出了框架语义结构抽取任务，包括目标词识别、框架识别和语义角色标注等任务。框架排歧作为框架识别的子任务也受到了广泛关注。

早期的框架排歧模型采用传统的机器学习方法，人工构建特征，使用条件随机场、最大熵、支持向量机等模型来建模。(Li et al., 2011) 使用窗口技术和BOW策略抽取了词包等若干特征，用最大熵模型建模，特征信息稀疏。(Li et al., 2013) 提出了特征模板的自动选择算法，通过打分机制将得分高的特征加入特征模板，使用最大熵模型来进行框架排歧。这些传统的机器学习算法，人工选择了大量特征导致空间维度过高，特征稀疏，费时费力。

随着深度学习的发展，近年来有许多研究采用神经网络模型来进行框架排歧。(Hermann et al., 2014) 使用WSABIE算法将目标词以及框架表示学习映射到同一空间，计算它们之间的距离进行框架识别。(Zhao et al., 2016) 提出了一种通用的框架识别模型，通过使用DNN架构来学习目标词的上下文特征进行框架识别，对于未登录词元和歧义词元的框架识别有了较好的泛化能力。(Zhang et al., 2017) 针对人工抽取特征使得特征空间维度过高和特征之间缺乏关联性的问题，在词语分布式表征的基础上提出了基于距离和词语相似度矩阵的框架排歧模型，证明词语分布式表征对框架排歧的有效性。(Botschen et al., 2017) 使用Word2Vec训练词向量表征上下文来进行框架识别。(Hou et al., 2020) 提出了一种基于hinge-loss的框架表示学习算法，通过计算目标词表示和框架表示之间的相似度来进行框架排歧，相较之前的工作有了明显的提升。(Guo, 2021) 将BERT与Bi-GRU结合起来编码上下文信息，使用注意力机制融入局部和全局信息来进行框架识别，在CFN和FrameNet上提高了框架识别的准确率。(Su et al., 2021) 通过融入框架关系和框架定义来进行框架识别。但以上工作都不是特定于目标词生成的向量表示，会保留与目标词无关的噪声信息；并且未充分考虑句法信息的重要性。在框架语义角色标注中，大多句法信息是目标词的语义角色，对于目标词的所属框架选择有着重要作用。(Li et al., 2010) 使用层次条件随机场 (T-CRF)，将框架识别视为依赖树结构上的标注任务。(Wang et al., 2013) 使用T-CRF模型建模，选取词、词性和不同类型的依存特征进行框架语义角色自动标注。但以上模型难以在不同的树结构上并行计算，计算效率低下。近年来，图卷积网络 (Graph Convolutional Networks, GCN) (Kipf and Welling, 2016) 的兴起为依存树的构建提供了新的思路。GCN可以有效存储任何结构的依赖树信息，并且能够并行计算，计算效率得到了很大的提升。(Zhang et al., 2018) 将BiLSTM和GCN结合起来编码句子中的句法信息，用来解决关系抽取任务，证明了GCN编码依存树的有效性。

因此本文提出了基于GCN和门机制的框架排歧模型生成特定于目标词的向量表示，并利用句法信息构造依存图，使用一种基于依存图的约束机制来约束模型学习，改进向量表示。相比基于上下文生成的向量表示，使用目标词附近的局部信息，句法信息对目标词更为重要。实验结果表明，模型在CFN和FrameNet框架排歧数据集上取得了一定的提升。

3 基于GCN和门机制的框架排歧模型

在框架排歧中，给定句子 $s : \{c_1, \dots, c_i, \dots, c_n\}$ 和目标词 c' (可能由多个字组成)，对于目标词 c' ，它所能激起的框架 $F = \{f_1, \dots, f_g\}$ 来自CFN的框架库，框架排歧任务就是在候选框架列表 $\{f_1, \dots, f_g\}$ 中为 c' 在当前语境句子 s 下找到最合适的框架，其形式化描述如公式 (1) 所示。

$$f = \arg \max_{f_i \in F} P(f_i | c', s) \quad (1)$$

本文提出了一种基于GCN和门机制的框架排歧模型，模型整体架构如图1所示，该模型是针对给定目标词和包含该目标词的句子，通过模型训练得到目标词的表征，进行框架排歧。模型的整体包括编码层、依存图抽取模块、融合门机制的图卷积网络层 (GGCN)、基于依存图的约束机制、分类层五个模块。本文通过预先训练的BERT来获取基于上下文的词表征，使用LTP进行句法分析，以此构建依存图，将图和词表征输入GCN来更新图节点的信息，为每

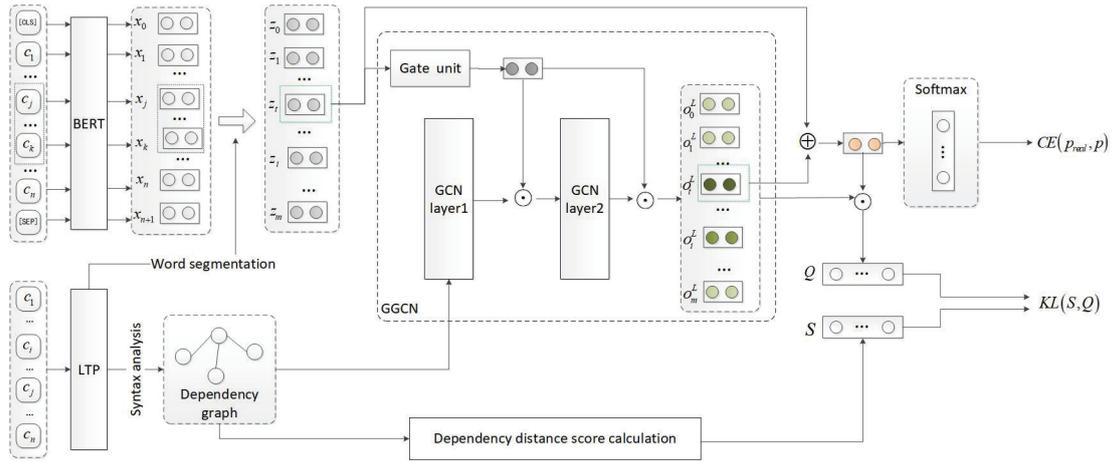


图 1: 基于GCN和门机制的框架排歧模型

个GCN层增加一个门，以过滤与目标词无关的信息；并依据依存图计算句子中每个词到目标词的距离，取其负值作为每个词基于句法的得分，输入模型，监督模型的训练，改进向量表示；将模型最终得到的目标词表征，通过softmax函数，得到一个向量表示，向量每个维度上的值为候选框架的概率分布，选择概率值最大的框架作为正确框架。

3.1 编码层

BERT是一个大规模的预训练模型，以无监督的方式对大规模未标记的语料库进行训练，能够大规模学习语言中隐含且丰富的文本语义。其体系结构是一种多层双向Transformer encoder，相比传统的Transformer拥有双向编码能力，可以更彻底的捕捉上下文信息获得动态词向量表示，具有更深的层数和并行性，进一步增加词向量模型泛化能力，充分提取到了字符级、词级、句子级和句间等特征。因此本文使用BERT作为编码层，对预训练的BERT进行了微调来适应框架排歧任务。

将“[CLS]+s:{ $c_1, \dots, c_i, \dots, c_n$ }+[SEP]”作为模型输入，编码层将输入中的每一个字符编码成字符嵌入 $E_{token}(c_i)$ 、分段嵌入 $E_{seg}(c_i)$ 和位置嵌入 $E_{pos}(c_i)$ 三个向量，将三个向量相加输入BERT预训练模型得到输入的BERT向量 $X \in R^{n \times d}$ ，如公式 (2)、(3) 所示。

$$E_i = E_{token}(c_i) + E_{seg}(c_i) + E_{pos}(c_i) \quad (2)$$

$$X = BERT(E_0, \dots, E_i, \dots, E_{n+1}) \quad (3)$$

Token Embedding: 模型通过查询向量表将输入中的每个字符转换为一维向量。

Segment Embedding: 框架排歧只输入模型一个句子，Segment Embedding全设为0。

Position Embedding: Transformers无法编码序列的顺序性，而文本中不同位置的字符携带的信息是不同的，应该用不同的向量表示，通过让BERT模型为每个位置学习一个位置嵌入来编码序列的顺序性。

本文采用对应于目标词token的BERT向量（如果有多个token，将其平均）作为目标词的表示 z_t ，如公式 (4) 所示。

$$z_t = avg(x_j, \dots, x_k) \quad (4)$$

3.2 依存图抽取模块

从图2中可以看出，与目标词有依存句法关系的词往往是目标词的框架元素，对目标词所属框架的确定起重要作用。目标词“炒”的直接依存信息主语 (SBV)、状语 (ADV) 和宾语 (VOB) 分别对应框架“烹饪”的框架元素“烹调者 (cook)”、“方式 (manr)”、“食物 (food)”等，但如果只考虑直接依存关系很可能会忽略目标词的一些有用的信息，例如“迎接我们的到来”也是“烹饪”的框架元素。因此本文采用多层GCN来聚合直接或间接依存信息。

本文直接使用哈工大的LTP工具包来进行分词和句法分析，利用得到的分词信息将BERT得到的字向量做平均得到词的表示作为GCN的节点输入。

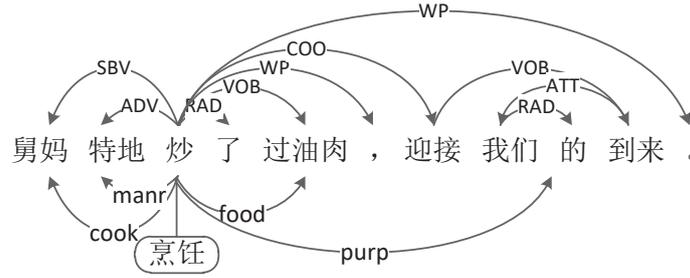


图 2: “炒”的依存关系和角色标注

3.3 融合门机制的图卷积网络层

从3.2中可知，框架排歧任务与目标词的依存句法信息密切相关。之前使用依存信息的框架排歧工作都是将依存特征通过拼接的方式来融入，特征不存在时用0向量表示，得到的特征向量维度高且稀疏。为了更好的融入目标词的依存信息，本文引入了图卷积网络（GCN），它是卷积神经网络（CNN）的一种改编，完善了CNN在非矩阵结构数据上的不适用性。GCN在进入下一层非线性变换之前，每个节点先将其自身的邻居节点的信息通过非线性方式聚合，相比BiLSTM和word2vec只利用了词周围的信息，GCN采用多层卷积，节点不仅利用了自身邻居节点的信息，同时也聚合了邻居的邻居信息，相比之下感受野更广，能够利用更大范围的信息。本文使用L层GCN来更新依存图的节点信息并抽出聚合依存句法信息的目标词表示。

首先定义无向图 $G = (V, E)$ 作为句子 s 的依存图，其中 $V = \{v_1, \dots, v_m\}$ 是图节点的集合，是将句子 s 经过LTP分词得到句子的分词结果。 E 是图边的集合， $(v_i, v_j) \in E$ 表示第 i 个词和第 j 个词之间存在有向句法弧，为了实现信息的反向传播，向 E 中添加一条与有向句法弧方向相反的边 (v_j, v_i) ，同时为了利用节点自身的信息，对所有的节点向 E 中添加一个自循环，即 (v_i, v_i) 。完成构图之后使用神经网络模型 $GCN(Z, A)$ 对图结构进行编码。首先获得节点的特征矩阵 Z 并计算图的邻接矩阵 A 。利用将经过BERT得到的字向量做平均作为图节点的特征矩阵 $Z \in R^{m \times d}$ ；通过 E 构建邻接矩阵 $A \in R^{m \times m}$ ， $A_{i,j} \in \{0, 1\}$ 表示第 i 个词和第 j 个词之间是否存在边。

GCN层的计算如公式（5）所示：

$$GCN(Z, A) = \hat{A} ReLU(\hat{A} Z W^{(l)}) W^{(l+1)} \quad (5)$$

其中 \hat{A} 是基于对角矩阵 D 的正则化邻接矩阵，计算如公式（6）所示：

$$\hat{A} = \tilde{D}^{-\frac{1}{2}} A \tilde{D}^{-\frac{1}{2}} \quad (6)$$

$W^{(l)}$ 为第 l 层的权值矩阵，用于将节点的特征表示映射到相应的隐层状态。 $W^{(l+1)}$ 为第 $l+1$ 层的权值矩阵，用于将节点的隐层表示映射为相应的输出。

由于GCN生成的隐层向量不是特定于目标词产生的，会导致隐层向量融入与目标词无关的噪声信息。而在框架排歧中，更关注与目标词相关的信息，因此本文引入门机制来进行特征过滤，利用目标词的嵌入表示 z_t 计算得到门向量 g^l ，通过元素乘积将门向量应用到GCN相应的隐层向量，得到过滤向量 o_i^l ，计算公式如下：

$$g^l = \sigma(W_g^l z_t) \quad (7)$$

$$o_i^l = g^l \odot h_i^l \quad (8)$$

其中， W_g^l 是GCN第 l 层的可学习参数， h_i^l 是GCN第 l 层第 i 个节点对应的隐层向量。

3.4 基于依存图的约束机制

在使用GCN融入句法信息时，更关注依存图中邻居节点的信息，虽然非相邻词可能不会直接为目标词带来有用的上下文信息，但本文认为其仍然可以为隐层向量的计算提供有用的训练信号。所以本文提出一种基于依存图的约束机制，根据依存图计算句子中每个词到目标词的距离，取其负值作为句子中每个词基于句法的得分 S ，并且本文认为框架排歧句子中每个词的

重要程度可以用其携带的有用信息来衡量，如果词 c_i 的过滤向量 o_i^L 与目标词的最终表示 z' 更相似，那么词 c_i 对于目标词更重要，使用公式 (9) 计算每个词相对于目标词基于模型的重要性得分 Q ，其中 W^z 和 W^o 是可训练参数。将 S 和 Q 分别经过softmax进行归一化。本文认为两个得分之间应该具有一致性，而KL散度是用来度量两个概率分布相似度的指标，因此通过计算两个得分之间的KL散度衡量两者之间的差异，并加入损失函数来最小化两者之间的差异，计算公式如式 (10) 所示。

$$q_i = \sigma(W^z z') \cdot \sigma(W^o o_i^L) \quad (9)$$

$$KL(S, Q) = - \sum_{i=1}^n s_i \frac{s_i}{q_i} \quad (10)$$

3.5 分类层

直接在GCN模块后计算损失、梯度下降，会导致梯度消失，所以本文使用线性插值法插值BERT得到的基于上下文的目标词表示 z_t 和GCN融合了句法信息的目标词表示 o_t^L 作为最终的目标词表征 z' ，如公式 (11) 所示，其中 μ 是权重系数。

$$z' = (1 - \mu) z_t + \mu o_t^L \quad (11)$$

将 z' 输入全连接层进行分类，通过softmax层计算各个候选框架概率值，如公式 (12) 所示。

$$p = \text{softmax}(z') \quad (12)$$

框架排歧为多分类任务，所以模型采用交叉熵损失函数作为分类损失，如式 (13) 所示，其中 p_{real} 表示真实样本类别分布。

$$CE(p_{real}, p) = - \sum p_{real} \log(p) \quad (13)$$

最后，本文采用式 (10) 和式 (13) 组合作为模型整体的损失函数来训练模型，其中 α 为权重系数，计算公式如下：

$$loss = CE(p_{real}, p) + \alpha KL(S, Q) \quad (14)$$

4 实验设计与分析

4.1 实验数据

本文使用的框架排歧数据来源于CFN数据库 (<http://sccfn.sxu.edu.cn/portal-en/frame.aspx>) 中抽取出来的88个有歧义的词元，共10012条数据，涉及到90个框架，训练集和测试集按8:1:1的比例分配，如表3所示。

同时为了验证模型的可行性和有效性，本文在FrameNet1.5和FrameNet1.7数据集上进行了实验。FrameNet1.5和FrameNet1.7数据分布如表3所示。

数据集	CFN	FN1.5	FN1.7
训练集	8001	16092	19152
验证集	1004	2197	2263
测试集	1007	4320	6698
词元数	88	2931	3416
框架数	90	704	796

表 3: 数据集分布

4.2 实验指标

本文采用准确率作为评价指标，计算如公式 (15) 所示。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

分子为排歧正确的例句数，分母为排歧的总例句数。

4.3 实验环境

本文实验环境如表4所示:

操作系统	Linux
CPU	Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz
GPU	Tesla P100-PCIE-16GB
Python	3.7.11
Pytorch	1.10.0

表 4: 实验环境

4.4 参数设置

在BERT中, 隐层维度768维, 最大序列长度采用512, batch_size设为4, learning_rate为1e-5, dropout设为0.1, GCN隐层维度为768。

4.5 实验结果与分析

本文分别在CFN和FrameNet数据集上进行了实验; 并对GCN的层数、 μ 和 α 进行了分析。

首先在CFN数据集上进行了实验, 本文设置了如下对比实验: (1) 使用BERT模型作为基线模型; (2) BiLSTM框架排歧模型; (3) (Hermann et al., 2014)、(Botschen et al., 2017)、(Hou et al., 2020)和(Guo, 2021)四组对比实验; (4) BERT+GCN (BGCN) 模型; (5) 使用门机制的BERT+GCN+GATE (BGCNG) 模型; (6) 使用约束机制的BERT+GCN+CON (BGCNC) 模型; (7) 使用门机制和约束机制的BERT+GCN+GATE+CON (BGCNGC) 模型。实验结果如表5所示。

Model	ACC
(Hermann et al., 2014)	63.48
(Botschen et al., 2017)	67.88
(Hou et al., 2020)	72.52
(Guo, 2021)	74.90
BiLSTM	71.70
BERT	73.98
BGCN	75.37
BGCNG	75.47
BGCNC	75.87
BGCNGC	75.97

表 5: CFN实验结果

表5中的实验结果显示, 本文提出的方法显著优于之前的方法, 相比 (Hermann et al., 2014)和 (Botschen et al., 2017)的方法分别提升了12.49%和8.09%, 相比 (Hou et al., 2020)提升了3.45%, 上述方法都是通过现有数据集学习到框架表示, 计算其与上下文表征的相似度, 数据集不均衡, 学到的框架表示较差, 并且未利用大规模预训练模型, 上下文特征抽取能力较弱。(Guo, 2021)的模型使用BERT作为编码器, BiGRU强化上下文语义表示, 用全局和局部注意力机制抽取目标词的全局信息和局部信息, BGCN相比 (Guo, 2021)提升了0.47%, 证明句法信息相比目标词局部信息和全局信息对于框架排歧来说更重要。从BERT和BiLSTM模型的实验结果可知, BERT相比BiLSTM可以更充分的利用上下文信息, 具有更强的信息表达能力。BGCNG通过增加门机制, 强化相对于目标词重要的信息, 降低冗余信息的影响, 一定程度上提升了框架排歧的准确率; BGCNC模型加入约束机制来监督模型学习, 充分利用目标词的句法信息, 相比BGCN, 准确率提升了0.5%; BGCNGC同时加入门机制和约束机制, 相比基线取得了较大的提升。以上结果证明了本文提出模型对汉语框架排歧的有效性。

为了验证模型的有效性和通用性，本文还在英文数据集FrameNet1.5和FrameNet1.7上进行了实验，并与之前的方法进行了对比，实验结果如表6所示。KGFI是 (Su et al., 2021)提出的模型，将框架定义和框架元素通过框架关系构图，利用GCN融入到框架表示当中来进行框架排歧，并采取了框架过滤的规则，本文不考虑框架过滤规则。从表中可知，本文提出的模型在数据集FrameNet1.5和FrameNet1.7上都有所提升，并且在利用外部知识的情况下在两个数据集上都超过了当前最新工作KGFI，在FrameNet1.5提升了0.33%，在FrameNet1.7提升了0.21%。

Model	FN1.5	FN1.7
(Hermann et al., 2014)	77.49	-
(Botschen et al., 2017)	81.21	-
KGFI	85.63	85.81
BERT	84.91	84.79
BGCN	85.32	85.01
BGCNG	85.63	85.45
BGCNC	85.77	85.85
BGCNGC	85.96	86.02

表 6: FrameNet实验结果

本文在数据集CFN、FrameNet1.5和FrameNet1.7探讨了GCN层数、 μ 、 α 对实验结果的影响。

依据图3可知，对于数据集CFN、FrameNet1.5和FrameNet1.7来说，GCN的层数为2时，取得的结果最好。经分析，当GCN层数为1时，目标词只聚合了自身邻居节点的信息，只利用了一阶依存信息，而对于目标词来说，有的二阶依存信息对其框架的选择也是必不可少的，相比之下两层GCN不仅聚合了自身的邻居信息，同时也融入了邻居的邻居信息，充分利用了二阶依存信息。但随着GCN层数的增加，在三个数据集上都表现出了下降的趋势，分析主要是因为GCN每次卷积都是节点和周围信息聚合的过程，随着GCN层数的增加，词节点聚合的信息越来越多，到最后会使得每个节点的嵌入表示变得非常相近，而框架排歧是针对句子中的某个词来排歧，并非在句子层面，给框架排歧带来了噪声干扰。

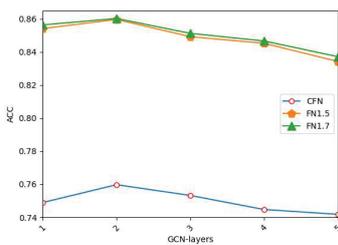


图 3: GCN层数的影响

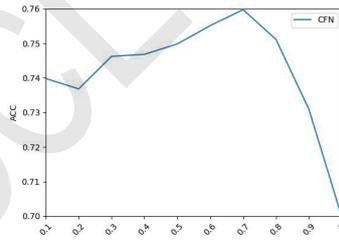


图 4: CFN上 μ 的影响

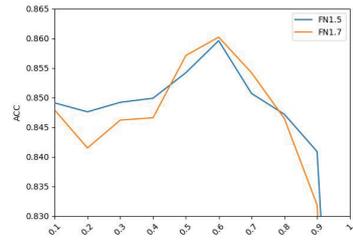


图 5: FrameNet上 μ 的影响

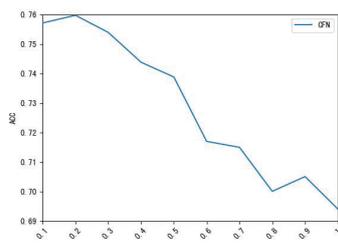


图 6: CFN上 α 的影响

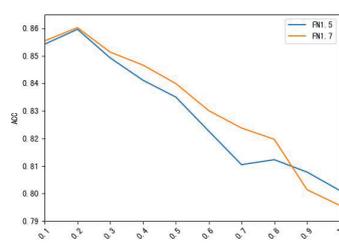


图 7: FrameNet上 α 的影响

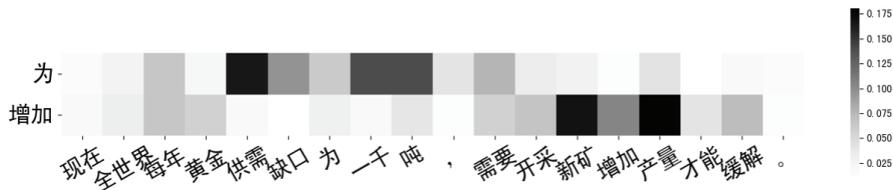


图 8: 门的可视化

本文设置了线性插值层来插值BERT和GCN的输出，插值参数 μ 控制BERT和GCN的权重，为了探究 μ 对整体模型的影响，在数据集CFN、FrameNet1.5和FrameNet1.7上对 μ 进行不同设置，结果如图4和图5所示。图4为10个epoch下CFN数据集上的 μ 参数分析，从图中可知，随着 μ 的增加，准确率越来越高，当 μ 为0.7时，模型效果达到最佳，表现效果好于仅使用BERT预测（ $\mu = 0$ ）和GCN预测（ $\mu = 1$ ）； $\mu > 0.7$ 时准确率开始下降，模型收敛相对较慢；当 $\mu = 1$ 时，GCN后直接计算损失、梯度下降，会出现梯度消失的问题，模型收敛速度变慢。图5为15个epoch下FrameNet数据集上的 μ 参数分析，从图中可知，在FrameNet数据集上，当 μ 为0.6时，模型效果达到最佳。从实验数据中可知，不同的数据集， μ 的最佳值不同。

图6、图7为 α 不同设置下的实验结果，从图中可知，当 α 取0.2时，准确率达到最优。当 α 取0.1到0.3时相对比较平稳，随着 α 的增大，准确率下降。

4.6 案例分析

本小节从框架排歧数据中选取了一条数据进行分析，通过对门权重进行可视化验证加入门机制的有效性。在标准化之后，绘制每个单词的值如图8所示。通过分析可知，目标词为“为”时，“供需”、“缺口”、“一千”、“吨”的权重较大，当目标词为“增加”时，“新矿”、“增加”、“产量”权重较大，由此可见门机制的加入，强化了对目标词重要的信息，降低了噪声信息的干扰。

5 总结

本文针对汉语进行了框架排歧研究，提出了一种基于GCN和门机制的框架排歧模型，相较于先前工作使用的词嵌入模型，该模型使用BERT增强了模型抽取特征的能力，可以获得更为丰富的语义信息，通过GCN将句法信息融入目标词表示中，使用门机制过滤与目标词无关的噪声信息，并提出一种基于依存图的约束机制来监督模型学习，改进向量表示。目前汉语框架排歧还存在一定的挑战，如现有框架排歧数据规模有限，并且存在数据不平衡问题，对于出现频率低的框架很难通过有监督的模型训练得到好的性能，如何更好的解决数据不均衡性，可否将元学习 (Kumar et al., 2019; Holla et al., 2020; Du et al., 2021; Chen et al., 2021) 方法应用到框架排歧上，以及模型对领域外的数据是否同样具有通用性还有待研究与解决。

参考文献

- Baker, Collin F and Fillmore, Charles J and Lowe, John B. 1998. *The berkeley framenet project*. COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics, pages:86-90.
- Baker, Collin F and Ellsworth, Michael and Erk, Katrin. 2007. *SemEval-2007 Task 19: Frame semantic structure extraction*. Proceedings of the 4th International Workshop on Semantic Evaluations, pages:99-104.
- Botschen, Teresa and Mousselly-Sergieh, Hatem and Gurevych, Iryna. 2017. *Prediction of frame-to-frame relations in the FrameNet hierarchy with frame embeddings*. Proceedings of the 2nd Workshop on Representation Learning for NLP, pages:146-156.
- Che, Wanxiang and Li, Zhenghua and Liu, Ting. 2010. *Ltp: A chinese language technology platform*. Coling 2010: Demonstrations, pages:13-16. Beijing, China.
- Chen, Howard and Xia, Mengzhou and Chen, Danqi. 2021. *Non-parametric few-shot learning for word sense disambiguation*. arXiv preprint arXiv:2104.12677.

- Devlin, Jacob and Chang, Ming-Wei and Lee, Kenton and Toutanova, Kristina. 2018. *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805, pages:4171-4186.
- Du, Yingjun and Holla, Nithin and Zhen, Xiantong and Snoek, Cees GM and Shutova, Ekaterina. 2021. *Meta-Learning with Variational Semantic Memory for Word Sense Disambiguation*. arXiv preprint arXiv:2106.02960.
- Fillmore, Charles J and others. 1976. *Frame semantics and the nature of language*. Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech, 280:20-32.
- 郭哲铭. 基于注意力机制的框架识别技术研究[D]. 山西大学, 2021.
- Hermann, Karl Moritz and Das, Dipanjan and Weston, Jason and Ganchev, Kuzman. 2014. *Semantic frame identification with distributed word representations*. Meeting of the Association for Computational Linguistics, pages:1448-1458. Baltimore, USA.
- 侯运瑶, 曹学飞, 崔军, 王瑞波, 李济洪, 李茹. 2020. 基于框架表示学习的汉语框架排歧. 计算机应用研究, 37(12):5.
- Holla, Nithin and Mishra, Pushkar and Yannakoudakis, Helen and Shutova, Ekaterina. 2020. *Learning to learn to disambiguate: Meta-learning for few-shot word sense disambiguation*. arXiv preprint arXiv:2004.14355.
- Kipf, Thomas N and Welling, Max. 2016. *Semi-supervised classification with graph convolutional network*. arXiv preprint arXiv:1609.02907.
- Kumar, Sawan and Jat, Sharmistha and Saxena, Karan and Talukdar, Partha. 2019. *Zero-shot word sense disambiguation using sense definition embeddings*. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistic, pages:5670-5681.
- 李茹. 汉语句子框架语义结构分析技术研究[D]. 山西大学, 2012.
- Li, Qimai and Han, Zhichao and Wu, Xiao-Ming. 2018. *Deeper insights into graph convolutional networks for semi-supervised learning*. Thirty-Second AAAI conference on artificial intelligence.
- 李济洪, 高亚慧, 王瑞波, 李国臣. 2011. 汉语框架自动识别中的歧义消解. 中文信息学报, 25(03):38-44.
- 李国臣, 张立凡, 李茹, 刘海静, 石佼. 2013. 基于词元语义特征的汉语框架排歧研究. 中文信息学报, 27(4):44-52.
- Li, Ru and Liu, Haijing and Li, Shuanghong. 2010. *Chinese frame identification using t-crf model*. Coling 2010: Posters, pages:674-682.
- 石佼, 李茹, 王智强. 2014. 汉语核心框架语义分析. 中文信息学报, 28(6):48-55.
- Su, Xuefeng and Li, Ru and Li, Xiaoli and Pan, Jeff Z and Zhang, Hu and Chai, Qinghua and Han, Xiaoqi. 2021. *A Knowledge-Guided Framework for Frame Identification*. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pages:5230-5240.
- 王智强, 李茹, 阴志洲, 刘海静, 李双红. 2013. 基于依存特征的汉语框架语义角色自动标注. 中文信息学报, 27(2):34-41.
- Wang, Xiaohui and Li, Ru and Wang, Zhiqiang and Chai, Qinghua and Han, Xiaoqi. 2020. 基于Self-Attention的句法感知汉语框架语义角色标注. Proceedings of the 19th Chinese National Conference on Computational Linguistics, pages:616-623.
- You, Liping and Liu, Kaiying. 2005. *Building chinese framenet database*. Natural Language Processing and Knowledge Engineering, pages:301-306.
- 赵红燕, 李茹, 张晟, 张力文. 2016. 基于DNN的汉语框架识别研究. 中文信息学报, 30(6):75-83.
- 张力文, 王瑞波, 李茹, 张晟. 2017. 基于词分布式表征的汉语框架排歧模型. 中文信息学报, 31(6):50-57.
- Zhang, Yuhao and Qi, Peng and Manning, Christopher D. 2018. *Graph convolution over pruned dependency trees improves relation extractio*. arXiv preprint arXiv:1809.10185.