

融合情感分析的隐式反问句识别模型

李翔, 刘承伟, 朱晓旭*

苏州大学计算机科学与技术学院, 苏州, 中国

{20195227066,20194027010}@stu.suda.edu.cn, xiaoxzhu@suda.edu.cn

摘要

反问是现代汉语中一种常用的修辞手法, 根据是否含有反问标记可分为显式反问句与隐式反问句。其中隐式反问句表达的情感更为丰富, 表现形式也十分复杂, 对隐式反问句的识别更具挑战性。本文首先扩充了汉语反问句语料库, 语料库规模达到 10,000 余句, 接着针对隐式反问句的特点, 提出了一种融合情感分析的隐式反问句识别模型。模型考虑了句子的语义信息, 上下文信息, 并借助情感分析任务辅助识别隐式反问句。实验结果表明, 本文提出的模型在隐式反问句识别任务上取得了良好的性能。

关键词: 隐式反问句识别; 情感捕捉; 上下文信息; 词性

Implicit Rhetorical Questions Recognition Model Combined with Sentiment Analysis

Xiang Li, Chengwei Liu, Xiaoxu Zhu*

School of Computer Science and Technology, Soochow University, Suzhou, China

{20195227066,20194027010}@stu.suda.edu.cn, xiaoxzhu@suda.edu.cn

Abstract

Rhetorical question is a commonly used rhetorical technique in modern Chinese. It can be divided into explicit rhetorical question and implicit rhetorical question according to whether it contains rhetorical question mark. The implicit rhetorical question expresses more emotion and is more complex in form. This paper first expands the Chinese rhetorical question corpus with a scale of more than 10,000 sentences, and then proposes an implicit rhetorical question recognition model, which is integrated with sentiment analysis. The model considers the semantic information and context information of the sentence, and assists in identifying implicit rhetorical questions with the help of sentiment analysis tasks. The experimental results show that the model proposed in this paper has achieved good performance on the task of implicit rhetorical question recognition.

Keywords: Implicit Rhetorical Question , Sentiment Capture , Contextual Information , POS

1 引言

反问句作为汉语一种极具表现力的表达方式, 具有巨大的艺术表现张力和极其灵活的表现形式。反问句通常以疑问的形式表达肯定或否定的情感, 在现代汉语中常常作为一种修辞方式。人们表达情感时通过使用反问句来加强了语气的渲染和强调, 从而表达出强度更大, 更为具体的情感。在反问句的划分上, 李翔 et al. (2021) 根据句子中是否含有反问标记 (刘芳, 2011) 将反问句分为显式反问句和隐式反问句。在例句 1 中, “难道” 一词属于较为典型的反问标记, 它的出现增加了整个句子的反诘度, 可以很容易地识别出例 1 为反问句。例 2 句中并没有可以作为判断依据的反问标记, 其利用三个疑问句表达说话人内心对国家队的失望, 属于“无疑而问”的反问句。

例句 1: 我不是足球迷, 也没看比赛 (没心情看), 为什么韩国人就行? 难道韩国就不是黄种人?

例句 2: 就这水平还叫国家队? 国家队就踢成这样? 你们脸呢?!

反问标记的作用在于其能够影响周围词语甚至整个句子的反诘度, 因此, 语言学研究者在研究反问句时多是以显式反问句为主, 从显式词语、特有句式结构等角度入手, 总结了许多具有代表性的反问句标记 (刘芳, 2011; 殷树林, 2007; 郭继懋, 1997)。与显式反问句相比, 隐式反问句的研究还处于探索阶段。一方面, 隐式反问句在句法结构上与一般疑问句没有明显区别, 可供研究者研究的语言特征较少, 另一方面, 隐式反问句由于不含反问标记, 在表达上更为复杂, 对其进行判断时需要借助相关的语用推理, 上下文信息等, 这也就导致隐式反问句的语料十分匮乏。

反问句往往可以表达出更细粒度的情感, 对其进行正确地识别, 可以改善情感分析的结果。同样, 情感分析也可以促进反问句的识别, 特别是对于缺乏明显特征的隐式反问句。例如例句 2 在句式上虽与疑问句并无不同, 但通过上下文可以知道说话人是以疑问句的形式表达负面情感, 其目的并不是提问, 因此例句 2 是无需回答的反问句。

针对隐式反问句识别研究所面临的挑战, 本文首先扩充了隐式反问句的语料, 接着提出了融合情感分析的隐式反问句识别模型 (Implicit Rhetorical Question Recognition Model Combined with Sentiment Analysis)。模型首先利用疑问词、副词、语气词等可以成为表达强烈反诘语气载体的特点引入词性特征以丰富句子的语义信息, 并捕捉句子内部的情感变化以确定不同上下文信息的权重, 再辅以情感分析技术识别隐式反问句, 最后引入标签正则化与稀疏正则化进一步提高模型性能。本文做出贡献总结如下:

(1) 利用李翔 et al. (2021) 提出的基于半监督学习和主动学习的半自动反问句语料收集方法扩充汉语反问句语料库, 语料库规模达到 10000 余句, 其中隐式反问句 5556 句。

(2) 本文提出融合情感分析的隐式反问句识别模型 IRQR。该模型深度挖掘了句子的语义信息, 情感信息以及上下文信息, 并利用这些信息对隐式反问句进行识别。

(3) 借助汉语反问句语料库进行了大量实验, 实验结果证明了 IRQR 模型在识别隐式反问句中的有效性。

©2021 中国计算语言学大会

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

基金项目: 国家自然科学基金 (61836007;61772354;61773276); 江苏高校优势学科建设工程资助项目

本文组织结构如下：第 2 节介绍关于反问句研究的相关工作。第 3 节详细阐述融合情感分析的隐式反问句识别模型。第 4 节介绍反问句识别实验。第 5 节给出本文的总结并展望未来工作。

2 相关工作

在语言学领域，语言学家们分别从不同的角度对反问句进行了研究。刘钦荣 (2004) 认为在深层语义结构中，由于反问句中不存在未知的语义成分，反问句不表达疑问而表达判断。刘彬 et al. (2017) 认为反问句的语用环境往往是“反通常性”的，说话人会在“疑有信无”、“疑是信非”等怀疑原则的指导下，使得反问句表达出否定的意思。他们认为听话人可以使用语言交际中的“合作原则”以及其下相关准则来理解反问句的意义。从反问句的交际作用角度出发，刘松江 (1993) 认为反问句带着十分浓厚的感情色彩，说话人通过使用反问句来宣泄自己的情感。殷树林 (2007) 探究了反问句与疑问句在句法结构上的异同，总结了 17 类反问句特有的句法结构。王敏 (2011) 从语用学角度出发，提出了分辨一句话是否为反问句的关键是语境的观点。在反问句的语义语用方面，郭继懋 (1997) 罗列出使用反问句所需要具备的三个语义语用条件。于天昱 (2018) 考察了反问句在对话中的位置和结构，认为反问句在使用上较为灵活，并多数反问句的出现伴随着用于表达情绪的语气词出现。

在应用计算机识别反问句方面，无论是在英文领域还是在中文领域，反问句识别的相关研究都还比较少。在英文领域，Bhattasali et al. (2015) 将 n-gram 语言模型和上下文信息作为反问句特征，利用 SVM 和贝叶斯分类器进行反问句识别，实验结果显示 SVM 分类器的实验效果最优，F1 值提升 10%，达到了 53.71%。Ranganath et al. (2016) 利用语言学特性为用户动机进行建模，挖掘用户所表达的隐式信息，识别社交媒体中的反问，并使用 Twitter 上发布的推文评估模型。Oraby et al. (2017) 则在前人工作的基础上，构建了一个基于 Debate Forums 和 Twitter 的英文反问句数据集，并利用 SVM 进行反问句识别实验，F1 值达到了 76%。同时，Oraby et al. (2017) 还对反问句的用途（讽刺与非讽刺）进行研究，将反问特征与上下文特征作为 SVM 的输入特征，对反问句的用途进行分类，在 Twitter 数据集上达到了 83% 的 F1 值。

在中文领域，文治 et al. (2019) 将反问标记作为识别反问句的重要特征，设计了一种融合句式结构的卷积神经网络反问句识别方法，该方法的 F1 值为 86.7%。李旻 et al. (2020) 则在文治 et al. (2019) 研究的基础上，提出了基于注意力机制 Bi-LSTM 的反问句语言特征自动抽取模型 AutoF，同时设计了一种基于语言特征自动获取的反问句识别方法，达到 90.7% 的精确率。李翔 et al. (2021) 利用半监督学习和主动学习，提出半自动的反问句语料收集方法，构建了面向新闻评论的汉语反问句语料库，并在反问句识别实验中引入了位置特征与句法路径特征，取得了 92.15% 的 F1 值。本文的工作则主要是针对隐式反问句，以填补这一方面研究的空白。在前人工作的基础上，本文进一步扩充了隐式反问句语料，将句子的词性特征，情感变化特征，上下文特征进行融合以强化句子的语言特征表示，同时以情感分析任务作为辅助任务，利用 GRU 神经网络模型对隐式反问句进行识别。

3 隐式反问句识别模型

本文提出了一种融合情感分析的隐式反问句识别模型 IRQR，模型的总体架构如图 1 所示。IRQR 模型主要包含以下几层：(1) 语义特征层 (Semantic layer)。该层将句子 S 中每个词语 $\{w_1, w_2, w_3, \dots, w_n\}$ 的词性 $\{pos_1, pos_2, pos_3, \dots, pos_n\}$ 作为衡量词语在句子中权重的标准，利用

注意力机制更准确地获取句子的语义特征。(2) 情感分析层 (Sentiment layer)。本文为句子标注情感标签，并在模型中引入情感分析任务来辅助反问句识别任务。在这一层中，本文将利用情感词典⁰抽取到的情感词 $\{Sen_{w_1}, Sen_{w_2}, Sen_{w_3}, \dots, Sen_{w_k}\}$ 作为主要的输入特征。(3) 上下文特征层 (Contex layer)。为最大限度获得数据的上下文信息，本文先对话料进行分句处理，得到 $\{Context_1, Context_2, Context_3, \dots, Context_m\}$ 。考虑到不同上下文信息对反问句的判断的作用不同，模型对每个 *Context* 所引起的情感变化进行捕捉，并利用根据情感变化的程度体现各 *Context* 的贡献。(4) 正则化层 (Regularization layer)。本文引入标签正则化与稀疏规则化来约束注意力权重分配，以帮助学习更好的句子表示形式。下面将针对模型进行详细地说明。

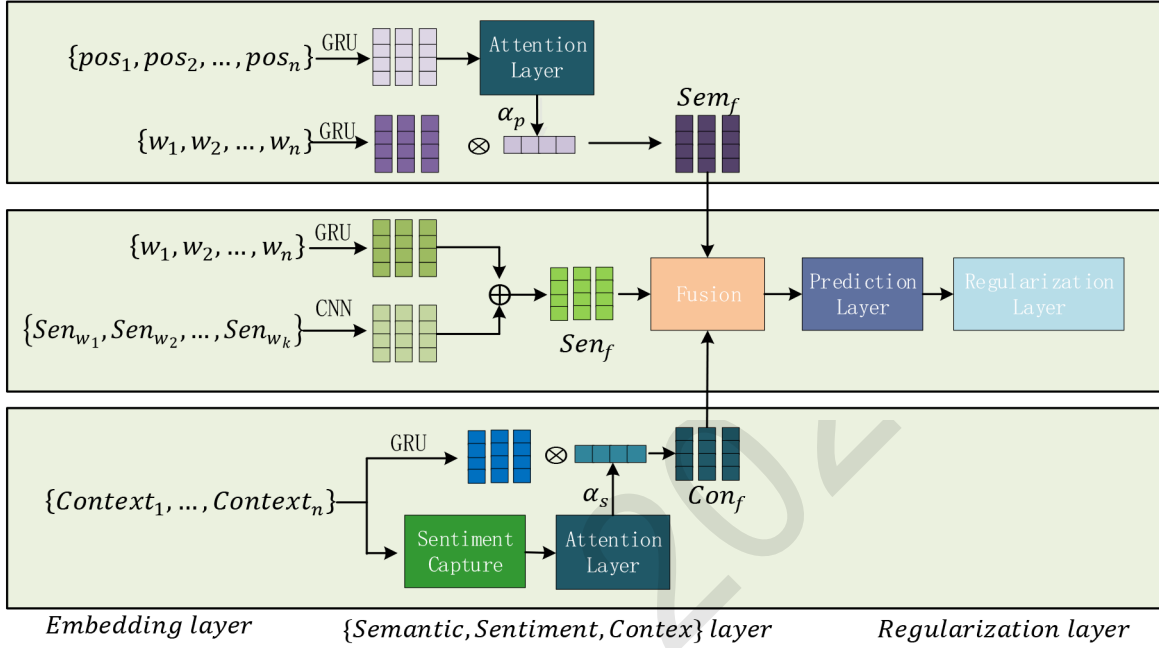


图 1: 模型结构图

3.1 语义特征层

由于 GRU(Cho et al., 2014) 网络可以捕获远间隔依赖关系的特性, 本文将其作为本层的编码器。首先, 将句子 $S = \{w_1, w_2, w_3, \dots, w_n\}$ 以及词性 $Pos = \{pos_1, pos_2, pos_3, \dots, pos_n\}$ 经过嵌入层之后分别输入到 GRU 网络中, 得到句子 S 、词性 Pos 的隐藏层表示 $h_w = \{h_{w_1}, h_{w_2}, h_{w_3}, \dots, h_{w_n}\}$, $h_p = \{h_{p_1}, h_{p_2}, h_{p_3}, \dots, h_{p_n}\}$, 词向量维度设置为 d_0 , 隐藏单元维度为 d_1 , $h_w \in R^{n*d_1}$, $h_p \in R^{n*d_1}$ 。通过注意力机制建立词语词性与任务类别标签的强相关性, 利用强相关性来体现词语的重要性, 进而判断该词性是否为反问句的特征。设 α_p^i 为句子 S 中词语 w_i 的词性 pos_i 的注意力权重, 其计算方式如公式 (1) 所示。

$$\alpha_p^i = softmax(tanh(W_p * h_{p_i} + b_p)) \quad (1)$$

其中, $W_p, b_p \in R^{l*2d_1}$, $\alpha_p^i \in R^{l*n}$, l 为任务类别标签数量。然后, 将词语与其对应的词性权重相融合得到句子中每个词语的表示。

$$w_{p_i} = \alpha_p^i h_{w_i} \quad (2)$$

⁰<http://114.215.64.60:8094/site2/index.php/zh>

$$S = \{w_{p_1}, w_{p_2}, w_{p_3} \dots w_{p_n}\} \quad (3)$$

为挖掘更深层次的语义特征, 本文再次使用 GRU 网络分别从正、反两个方向处理 S , 从而获得句子的正向表示 \vec{h} 以及反向表示 \overleftarrow{h} , 将二者进行拼接产生句子的语义特征 Sem_f , 如公式 (4) 所示。

$$Sem_f = \vec{h} \oplus \overleftarrow{h} \quad (4)$$

其中, $\vec{h}, \overleftarrow{h} \in R^{d_1}$, $Sem_f \in R^{2d_1}$ 。

3.2 情感分析层

与其他句型不同, 反问句作为表达情感的载体, 往往能够表达出更为具体、强烈的情感, 不同句子所表达的情感与其使用的句型有着一定的联系。反问句表达的情感往往与其字面意思相反。语料中反问句多表达负面情绪, 非反问句则以中性情感为主。本文认为模型在学习过程中如能有效地识别句子所表达的情感将有助于提升反问句识别的性能, 因此, 在这一层中本文将情感分析作为模型学习的目标, 得到的情感特征则为隐式反问句的识别提供情感信息。

首先句子 S 经过映射后获得相应的矩阵表示, 接着将句子的矩阵表示输入到 GRU 网络中学习情感的句子表示 h_{sen} , 具体计算方式可参考公式 (4)。值得注意的是, 情感分析层的 GRU 网络与语义特征层的 GRU 网络并不共享隐藏单元。

在情感分析任务的研究过程中, 研究者已经证明文本中的情感词能够从某种程度上体现作者想要表达的情感, 因此本文在将情感词作为主要的情感特征。根据 CNN(LeCun et al., 1989) 网络能够有效捕捉局部特征的特点, 本文将抽取出的情感词 $\{Sen_{w_1}, Sen_{w_2}, Sen_{w_3}, \dots, Sen_{w_k}\}$ 经过映射后输入到 CNN 网络中利用 t 个不同大小的卷积核对其进行卷积, 再经过池化, 激活, 拼接等操作最终得到新的特征 h_w , 如公式 (5)、(6)、(7) 所示。

$$C = Conv(\{Sen_{w_1}, Sen_{w_2}, Sen_{w_3}, \dots, Sen_{w_k}\}) \quad (5)$$

$$h = Relu(Maxpool(C)) \quad (6)$$

$$h_w = [h_1, h_2, h_3, \dots, h_t] \quad (7)$$

最后, 为加强这些情感词对情感分析的影响, 本文将情感的句子表示 h_{sen} 与基于情感词的情感特征 h_w 相融合, 得到新的情感特征 Sen_f , 计算方式如公式 (8)。

$$Sen_f = h_{sen} + h_w \quad (8)$$

3.3 上下文特征层

隐式反问句的特点在于其没有明显的反问特征, 识别的过程中高度依赖语境, 因此模型要尽可能地获取句子的上下文信息以帮助隐式反问句的识别。对于上下文文本 $\{Context_1, Context_2, Context_3, \dots, Context_m\}$, 模型首先借助 GRU 网络获得 $Context_i$ 的句子表示 h_{con_i} , 计算方式可参考公式 (4)。反问句往往能够带来较为明显的情感波动, 如果模型能够捕捉到上下文文本之间的情感变化就可以判断出不同上下文对任务的重要性。首先, 模型要获得 $Context_i$ 的情感极性:

$$p_{s_i} = softmax(W_{con} * h_{con_i} + b_{con}) \quad (9)$$

其中, $W_{con}, b_{con} \in R^{c*2d_1}$, c 为情感类别。对于 $Context_i$ 与 $Context_j$ 之间的情感变化, 本文采用公式 (10)、(11) 计算:

$$\Delta_{ij} = D_{KL}(p_{s_i} || p_{s_j}) \quad (10)$$

$$D_{KL} = \frac{1}{2} \sum_l^c p(l) \log q(l) + q(l) \log p(l) \quad (11)$$

其中, D_{KL} 表示对称 KL 散度, $p(l), q(l)$ 为情感标签分布。

接着利用情感偏移 $\Delta_{i,ij}$ 来确定每个 $Context$ 在文本中的权重, 其注意力权重计算方式如下:

$$\alpha_s = softmax(tanh([\Delta_{1,12}, \Delta_{12,123}, \dots, \Delta_{1\dots m-1,1\dots m}])) \quad (12)$$

得到 $Context$ 的权重之后, 根据公式 (13) 计算最终的上下文特征。

$$Con_f = \sum \alpha_{s_i} h_{con_i} \quad (13)$$

3.4 正则化层

Hu et al. (2018) 在多方面情感分析任务中引入正则化限制模型中的注意力分布, 类似地, 本文在这一层将标签正则化和稀疏正则化分别作用于任务标签集的概率分布和上下文权重以提高模型的性能。

如果模型对某一样本是否为隐式反问句的预测概率为 58%, 那么本文认为模型的鲁棒性较差, 对该样本并不是十分“确定”, 模型仍然有可以学习的空间。本文在模型输出端得到的任务标签集的概率分布 p 上引入标签正则化 R_l 来体现这种“不确定”, 并将 R_l 定义成如下公式:

$$R_l = |\sum_i^l p_i^2 - 1| \quad (14)$$

其中, $\sum_i^l p_i^2 = 1$ 且 $p_i > 0$ 。最小化 R_l 与最小化 p 的熵具有同样的效果, 也就是使模型的预测结果更加准确。

稀疏正则化 R_s 主要针对的是每个 $Context$ 的注意力权重 α_s 。每个 $Context$ 对句子的作用不尽相同, 如果能够将注意力权重的分布限制在更有意义的 $Context$ 上, 就可以获得对任务更有意义的上下文信息, 因此, 要尽可能地使 α_s 变得稀疏。稀疏正则化 R_s 在实现上与 R_l 类似, 这里不再赘述。

3.5 预测层

隐式反问句预测 将情感特征 Sen_f 以及上下文特征 Con_f 拼接在一起, 并进行最大池化操作, 再与语义特征 Sen_f 相融合, 得到最终的句子表示, 计算方式如以下公式:

$$Sen_{con} = Sen_f \oplus Con_f \quad (15)$$

$$Sen_{con} = Maxpool(Sen_{con}) \quad (16)$$

$$H = Relu(Sen_f + Sen_{con}) \quad (17)$$

最后将 H 输入 softmax 中, 判断是否属于隐式反问句。计算方式如公式 (18), 其中 $W_{rq} \in R^{l*2d_1}$, $b_{rq} \in R^l$ 。

$$Y_{rq} = softmax(W_{rq} * H + b_{rq}) \quad (18)$$

情感分析预测 将情感特征 Sen_f 输入 softmax, 判断句子情感类别。计算方式如以下公式 (19), 其中 $W_{sa} \in R^{c \times 2d_1}$, $b_{sa} \in R^c$ 。

$$Y_{sa} = \text{softmax}(W_{sa} * H + b_{sa}) \quad (19)$$

两个任务的在训练过程中均采用交叉熵损失函数, 隐式反问句识别任务的 Loss 记作 L_{rq} 情感分析任务的 Loss 记作 L_{sa} , 模型整体 Loss 如公式 (20) 所示。本文共同为两个任务训练我们的模型, 通过最小化损失函数来训练模型中的参数。

$$Loss = L_{rq} + \frac{L_{sa}}{c} + \lambda(R_l + R_s) \quad (20)$$

其中, λ 表示用于调整从正则化损失到总损失影响的超参数。

4 隐式反问句识别实验

本节首先介绍隐式反问句识别实验的数据来源, 然后给出实验设置, 其中包括衡量性能指标、使用工具、实验参数的设置等, 接着介绍用于与本文提出的模型相对比的基准系统, 最后给出模型在语料库上的实验结果与分析。

4.1 实验数据与设置

本文实验数据采用李翔 et al. (2021) 构建的面向新闻评论的汉语反问句语料库, 并在此基础上, 根据其提出的基于半监督学习和主动学习的半自动反问句语料收集方法利用新闻评论数据对语料库进行扩充, 最终获得反问句 10500 句, 其中隐式反问句占比 52.92%。语料库相关数据统计如表 1 所示。

	句子数	句子平均长度 (以字为单位)
反问句	10500	49.19
非反问句	10144	46.91
显式反问句	4944	49.84
隐式反问句	5556	48.31

表 1: 语料库相关数据统计

实验过程中, 所有数据均经过去噪、去重和分词等预处理操作, 使用 HanLP¹对文本词性进行标注, 利用腾讯 AI 开放平台²为文本标注情感标签。本文将数据集的 80% 作为训练集, 10% 作为验证集, 10% 作为测试集, 采用精确率 (Precision)、召回率 (Recall) 和 F1 值 (F1-measure) 作为实验结果的评价指标, 并且所有实验均进行十倍交叉验证, 取平均值作为实验结果。

对于本文提出的 IRQR 模型, 其中的超参数设置为 $d_0 = 300$, $d_1 = 256$, batch size 为 50, dropout 为 0.5, 学习率为 $1e-3$ 。借助验证集, 本文将 λ 设置为 0.1。模型采用 Adam Optimization 优化算法更新模型参数, 共进行 10 次训练, 并保存在验证集上表现最好的模型。

¹<https://github.com/hankcs/HanLP>

²<https://ai.qq.com/>

4.2 对比实验

为对比不同模型的隐式反问句识别性能，本文选取以下模型作为基准系统并获得实验结果如表 2 所示。

- 1)ATT_LSTM: 通过词向量表将句子转换成相应的矩阵表示，再输入到 Bi-LSTM(Hochreiter and Schmidhuber, 1997) 网络中，然后利用注意力机制生成句子的加权表示，最后输入到 softmax 中进行隐式反问句识别。
- 2)GRU: GRU 是 RNN 网络效果较好的一种变体，和 LSTM 网络一样，也是为了解决长期记忆和反向传播中的梯度等问题而提出来的，但 GRU 模型中只引入两个门：分别是更新门和重置门，因此其结构比 LSTM 网络更为简单。
- 3)Transformer: Transformer(Vaswani et al., 2017) 摒弃了 RNN 的递归结构，采用自注意力(Self-Attention) 机制，克服了 RNN 训练速度缓慢的缺点，使得模型可以并行化训练，而且能够拥有全局信息。模型超参数设置：隐藏状态维度为 1024，迭代次数为 30，attention heads 为 5，学习率为 5e-4。
- 4)SPD_GRU: 李翔 et al. (2021) 利用反问句语料库，以句子的语义特征作为基本特征，将位置特征、句法路径特征与其相融合共同作为句子表示，并输入至神经网络模型进行反问句识别。本文以 GRU 网络为基础，将该方法应用于单独的隐式反问句识别。
- 5)Sarcasm_Mutimask: Majumder et al. (2019) 着眼于情感分析任务与讽刺识别任务之间的关联性，提出一种基于多任务学习的情感分析与讽刺识别模型。该模型中情感分析模块与讽刺识别模块相互辅助，证明了基于多任务学习的方法明显优于独立情感和讽刺分类器，同时，证明了情感分析任务与讽刺识别任务是相关的。
- 6)Sarcasm_Bert: Srivastava et al. (2020) 提出了一种新型的分层 Bert(Devlin et al., 2018) 模型用于讽刺识别。模型首先从句子中提取局部特征，接着使用卷积模块总结上下文中的所有句子。然后，将上下文传递给递归层以从输入中提取时间特征。最后对时间特征与输入的句子表示进行卷积，以检测句子是否具有讽刺意味。

模型	Precision%	Recall%	F1- measure%
ATT_LSTM	85.88	86.43	86.08
GRU	86.12	88.10	87.06
Transformer	85.36	85.39	85.36
SPD_GRU	87.01	87.58	87.21
Sarcasm_Mutimask	84.40	88.88	86.49
Sarcasm_Bert	87.62	88.89	88.25
IRQR	89.70	90.06	89.86

表 2: 各个模型的实验性能比较

从表 2 可以看出，本文提出的 IRQR 模型在性能上都要优于基准模型。Sarcasm_Bert 模型利用 Bert 模型强大的语义表示能力对句子进行编码，并利用卷积模块，递归模块提取句子特征，使得 F1 值达到了 88.25%。与 Sarcasm_Bert 模型相比，IRQR 模型则是从已有数据出发，针对隐式反问句特点搭建，取得了更好的效果，在精确率上提高了 2.12%，召回率提高了 1.17%，F1

值提高了 1.61%。

从实验结果上来看,只使用 LSTM、GRU、Transformer 网络来对隐式反问句进行识别并不能取得十分理想的效果,这是因为神经网络模型往往很难挖掘深层次的语言特征,对于像隐式反问句这种较为复杂的修辞手法,单一的神经网络模型只能学习简单的句子表示,因此对于隐式反问句的识别仍然需要借助有效的语言特征。

SPD_GRU、Sarcasm_Mutimask、IRQR 模型均以 GRU 网络为基础,但从 F1 值来看,Sarcasm_Mutimask 的实验性能较差,只达到了 86.25%。Sarcasm_Mutimask 模型与 IRQR 模型都利用情感分析任务辅助隐式反问句识别任务,但 Sarcasm_Mutimask 模型只能获取句子的情感特征,忽略了上下文信息对隐式反问句识别的重要性。SPD_GRU 模型通过句法路径特征以及位置特征来获取有效的上下文信息,但 SPD_GRU 模型在提取句法路径特征与位置特征时是以句子中动词为核心,这样就导致其获取的上下文信息只围绕核心动词,无法获得完整的语境。本文提出的模型则较为充分地利用了每一个 *Context* 的语义特征,并通过捕捉其情感变化确定其权重,这样使得模型拥有了充分的上下文信息用于识别隐式反问句。

4.3 消融实验

为了研究模型中不同部分对实验结果的影响,本文对 IRQR 模型进行了以下消融研究:(1) IRQR-C: IRQR 模型去掉上下文特征层;(2) IRQR-S: IRQR 模型去掉情感分析层;(3) IRQR-R: IRQR 模型去掉正则化层;(4) IRQR-C-S: 上下文特征层中不捕捉句子内部情感变化而直接利用上下文特征;(5) IRQR-CP: IRQR 模型上下文特征层不捕捉句子内部情感变化且语义特征层不引入词性特征;(6) IRQR-P: 语义特征层不引入词性特征。消融实验结果如表 3 所示。

模型	Precision%	Recall%	F1- measure%
IRQR-C	88.66	89.27	88.90
IRQR-S	88.31	89.79	89.01
IRQR-R	89.53	89.31	89.38
IRQR-CS	89.48	88.73	89.05
IRQR-CP	89.44	89.39	89.39
IRQR-P	89.60	90.01	89.78
IRQR	89.70	90.06	89.86

表 3: IRQR 模型消融实验结果

从消融实验的结果可以看出,即使 IRQR 模型去掉某一部分,其识别隐式反问句的结果也要优于基准系统中的 Sarcasm_Bert 模型,F1 值的波动幅度小于 1%,这表明本文提出的 IRQR 模型具有较好的鲁棒性。

当 IRQR 模型去掉上下文特征层时,即模型 IRQR-C,实验效果下降最为明显,精确率下降 1.04%,召回率下降 0.79%,F1 值下降 0.96%,充分说明上下文特征层是整个模型的关键,而如何提取上下文信息也是能否正确识别隐式反问句的关键。如果去掉模型的情感分析层,即不再使用情感分析任务辅助隐式反问句识别任务,精确率下降显著,这证明了情感分析任务与反问句识别任务之间的关联性。利用反问句与非反问句表达不同情感这一点,能够对识别隐式反问

句起着促进作用。正则化层对实验结果的影响主要体现在召回率上，去掉正则化层的 IRQR 召回率下降 0.75%。

如果保留上下文特征层但不捕捉句子内部情感变化 (IRQR-CS)，实验结果与 IRQR-C 相比有所上升，但其精确率与 IRQR 模型相比仍下降 0.22%，因此，捕捉句子内部的情感变化可以获取更有效的上下文特征。本文将词性特征作为隐式反问句的语言特征，并用词性特征来帮助模型获取更深层次的语义特征，实验结果表明不使用词性特征的情况下，模型的性能有所下降。

4.4 错误分析

为了研究 IRQR 模型的不足，本文选取 3 个错误案例进行分析

例 3：球员是不是他选的？战术是不是他定的？训练是不是他带的？临场是不是他指挥的？那输球了不怪他？

例 3 是由连续的选择问句构成的反问句，形式上与一般疑问句相同，句子内部没有明显情感变化。对于这样的句子，模型很难为其调整上下文信息的注意力权重，也就无法将注意力集重到分句“那输球了不怪他？”。

例 4：至少最后三个球都是门前有效解围啊，前面国足明明还在战斗，后面就有这些妖魔鬼怪在那里挖墙脚，真让人寒心！此外，大家要明白，剪辑的镜头是可以有诱导性的。把失误剪辑来毁人，把成功剪辑来推销，当大家没有判断力？

模型虽然对评论文本进行分句处理以减少冗余的语境信息对任务的影响，但当上下文信息过长而反问描述较为简短时，模型仍会忽略反问描述对句子的影响，造成误判。事实上，例 3 与例 4 都属于事实型的文本描述，对于这一类反问句的识别往往需要一定的逻辑判断，虽然已经为模型提供了语义特征，情感特征以及上下文特征，但要想正确识别这类反问句，模型仍需要获得关联性较强的信息。

例 5：说的很对！从第一场就这样认为。其他多数人就像没头的苍蝇，瞎踢，踢到哪里算哪里。可惜了蒿俊闵的努力！

模型利用词性来确定每个词语的注意力权重，然而由于词语词义的多样性，模型对于不常出现的词语词义并不能很好地理解与学习，将只注意力集中到了隐式反问句中经常出现的词语上，导致分类错误。例 5 中的“哪里”虽然属于疑问代词，但并不表示疑问，而是泛指任何场所。

5 总结与展望

针对反问句研究的现状，本文提出了隐式反问句识别任务。本文首先扩充了面向新闻评论的汉语反问句语料库，语料库规模达到了 10000 余句，接着本文提出了融合情感分析的隐式反问句识别模型 IRQR。模型利用词性特征丰富句子的语义信息，捕捉句子内部的情感变化以确定不同上下文信息的权重，并借助情感分析任务识别隐式反问句，同时引入标签正则化与稀疏正则化增强模型鲁棒性。实验结果表明，本文提出的 IRQR 模型在隐式反问句识别任务上取得了良好的性能。

在对模型进行错误分析时，本文发现模型虽然可以有效捕捉上下文信息，但是对于需要借助逻辑事理进行语用推理才能理解的上下文信息，模型仍然无法有效地挖掘。在接下来的工作中，本文将针对发现的问题展开进一步的研究。

参考文献

- 刘彬, 袁毓林, et al. 2017. 反问句否定意义的形成与识解机制. *语文研究*, (2017年04):12-19.
- 于天昱. 2018. 反问句在话语进程中的作用. *《广西师范大学学报》(哲学社会科学版)*, 54(3):96-102.
- 刘芳. 2011. 现代汉语反问句的标记研究. Ph.D. thesis, 沈阳师范大学.
- 刘钦荣. 2004. 反问句的句法, 语义, 语用分析. Ph.D. thesis.
- 刘松江. 1993. 反问句的交际作用. *语言教学与研究*, 2(47):13.
- Shohini Bhattachali, Jeremy Cytryn, Elana Feldman, and Joonsuk Park. 2015. Automatic identification of rhetorical questions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 743-749.
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- 王敏. 2011. 从语境角度看反问句的识别, 理解与翻译——以《坛经》为例. *开封教育学院学报*, 2.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- 郭继懋. 1997. 反问句的语义语用特点. *中国语文*, 2:111-121.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735-1780.
- Mengting Hu, Shiwan Zhao, Li Zhang, Keke Cai, Zhong Su, Renhong Cheng, and Xiaowei Shen. 2018. Can: Constrained attention networks for multi-aspect sentiment analysis. *arXiv preprint arXiv:1812.10735*.
- Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541-551.
- Navonil Majumder, Soujanya Poria, Haiyun Peng, Niyati Chhaya, Erik Cambria, and Alexander Gelbukh. 2019. Sentiment and sarcasm classification with multitask learning. *IEEE Intelligent Systems*, 34(3):38-43.
- Shereen Oraby, Vrindavan Harrison, Amita Misra, Ellen Riloff, and Marilyn Walker. 2017. Are you serious?: Rhetorical questions and sarcasm in social media dialog. *arXiv preprint arXiv:1709.05305*.
- Suhas Ranganath, Xia Hu, Jiliang Tang, Suhang Wang, and Huan Liu. 2016. Identifying rhetorical questions in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10.
- Himani Srivastava, Vaibhav Varshney, Surabhi Kumari, and Saurabh Srivastava. 2020. A novel hierarchical bert architecture for sarcasm detection. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 93-97.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762*.
- 李翔, 朱晓旭, et al. 2021. 面向新闻评论的汉语反问句语料库构建. *山西大学学报(自然科学版)*, 44(03):1-9.
- 文治, 李旻, 王素格, 廖健, and 陈鑫. 2019. 融合反问特征的卷积神经网络的中文反问句识别. *中文信息学报*, 33(1):68-76.
- 李旻, 吴卓嘉, 王素格, and 梁吉业. 2020. 基于语言特征自动获取的反问句识别方法. *中文信息学报*, 34(2):96-104.
- 殷树林. 2007. 现代汉语反问句特有的句法结构. *湖南科技大学学报(社会科学版)*, 10(3):101-105.