# 将语义信息融入 Retro-Reader 的粗略阅读模块以实现越南语机器阅读理解

**Hang Thi-Thu Le[1, 2], Viet-Duc Ho[1, 2], Duc-Vu Nguyen[*,1,2], Ngan Luu-Thuy Nguyen[1, 2]**
越南胡志明市信息技术大学[1]越南胡志明市越南国家大学
{18520274,18520610}@gm.uit.edu.vn {vund,ngannlt}@uit.edu.vn

arXiv:2301.00429v1 [cs.CL] 1 Jan 2023

摘要—**近年来，机器阅读理解已成为自然语言处理领域最先进、最热门的研究主题之一。可回答性问题的分类是机器阅读理解中一项相当重要的子任务；然而，相关研究并不多。Retro-Reader 是一项有效解决这一问题的研究。然而，大多数传统机器阅读理解模型，尤其是 Retro-Reader 的编码器，尚未能充分利用上下文的上下文语义信息。受 SemBERT 启发，我们使用语义角色标注（SRL）任务中的语义角色标签，为 mBERT、XLM-R、PhoBERT 等预训练语言模型添加语义信息。本实验旨在比较语义信息对越南语机器阅读理解中可回答性分类的影响。此外，我们希望通过本实验增强 Retro-Reader模型粗略阅读模块的编码器。增强后的、融入语义信息的 Retro-Reader模型编码器首次应用于越南语机器阅读理解任务，并取得了积极成果。**

索引词—**机器阅读理解、语义角色标注、SemBERT、自然语言处理、自然语言理解**

## I. 引言

机器阅读理解（MRC）是自然语言处理中的一项具有挑战性的任务；该任务要求模型根据特定段落回答问题 [1], [2]。为了准确回答 MRC 任务中的问题，模型需要具备较强的语言阅读和理解能力。近年来，MRC 模型不仅必须能够正确回答 [3], [4] 问题，还必须能够区分不可回答问题，以避免给出不恰当的答案[5]。

确定机器阅读理解任务中的作答能力至关重要；早期研究已在英语领域部分解决了这一问题 [6], [7]。通常，Zhang 等人提出的 Retro-Reader 改进模型 [8] 取得了优于以往研究的最先进的结果。

据 [8], 目前所知，标准阅读系统由两个模块组成：1）具有强大语言模型的编码器，以及 2）根据机器阅读理解（MRC）任务特点精心设计的解码器。回顾式阅读器（Retro-Reader）设计了一种阅读理解模型，通过整合阅读与验证策略的两个阶段，增强了分类问题的处理能力。然而，该模型过于关注解码器，而对编码器方面关注不足。

预训练语言模型（PrLMs），如 ELMo [9],GPT [10], BERT [11], RoBERTa [12], ALBERT [13],，在各种自然语言处理任务中取得了成功，广为人知，并发挥着强大编码器的作用。然而，现有语言模型只专注于利用纯上下文特征，很少考虑显式的上下文语义线索。这可能是当前语言表示模型的缺点之一。

Zhang 等人 [14] 将 BERT 的纯上下文表示与来自语义角色标签的显式语义相结合，以获得更深层次的意义表示，其结果优于常规 BERT，达到了最先进的水平。这表明，为预训练语言模型提供显式语义是必要的。

许多英语研究利用语义角色标签中的语义信息来改进自然语言理解（NLU）任务。然而，关于越南语语义角色标注任务的研究寥寥无几，现有研究主要致力于开发语义标注系统 [15]–[17] ，并未利用语义角色标签中的语义来处理其他任务。这正是我们开展本研究的动机。

受 SemBERT [14] 的启发，并以改进 Retro-Reader模型的编码器为目标，我们提出了一种融入语义信息的 Retro-Reader模型，以比较语义对越南语机器阅读理解可回答性分类的影响。因此，我们将语义信息整合到 Retro-Reader 分类器模块的编码器中。

本研究的主要贡献如下：

- 比较语义角色标注任务所提供的语义对越南语机器阅读理解问题可回答性分类的影响。
- Retro-Reader 模型粗略阅读模块的改进编码器。

*通讯作者

- 展示语义角色标签在越南语中的重要性。

## II. 相关工作

### A. 机器阅读理解

机器阅读理解的起源可追溯至20世纪70年代，当时研究人员开始认识到文本理解在人工智能中的重要性。W Lehnert 于 1977 年提出了首个阅读理解系统 QUALM [18]，为后续机器阅读理解研究奠定了基础。

到 2013 年，机器阅读理解已被明确描述并作为一项监督学习任务来处理：给定一段文本和一个相关问题，模型应按照要求的格式给出答案。英语机器阅读理解任务长期以来一直在发展，并早早推出了数据集[5], [19]。相比之下，越南语机器阅读理解问题直到最近才开始受到关注，并出现了若干数据集，例如 UIT-ViQuAD1.0 [20], ViMMRC [21], UIT-ViNewsQA [22]。与此同时，也有许多关于越南语机器阅读理解系统的研究 [23], [24]。上述机器阅读理解系统面临的一个问题是，它们的设计假设所有问题都可以回答，但现实生活中并非总是如此。因此，研究人员后来在机器阅读理解数据集中加入了不可回答问题 [5], [25], [26]。这对后续机器阅读理解模型构成了重大挑战；成功的机器阅读理解模型必须能够妥善处理两个方面：为可回答问题提供准确答案，并识别问题的可回答性。

在自动阅读理解中，区分不可回答问题是一项相对重要的任务，但以往研究并未关注这一问题，或仅采用简单的方法。大多数问题的可回答性分类任务都是通过答案片段预测进行训练的。Liu 等人 [27]在上下文中添加了一个 [分类标记] 词元，并在其机器阅读理解模型中使用了额外的简单分类器层；Back 等人 [7] 则依靠注意力机制识别问题中导致其无法回答的词语，从而为问题计算得分，并根据该得分判断其可回答性。

与以往研究不同，Retro-Reader [8]模型采用了一种全新的方法，包含两个策略阶段：阅读与验证。该模型在英语机器阅读理解任务上取得了最先进的结果。本研究将改进 Retro-Reader 模型，并将其应用于越南语机器阅读理解。

### B. 显式上下文语义

语义角色标注任务旨在通过分析自然语言文档的浅层语义，识别句子的谓词-论元结构。它有助于我们找到以下问题的答案：“谁？”“做了什么？”“为谁？”“用什么？”“在哪里？”以及“何时？” [28], [29]。巧合的是，这些正是一般的机器阅读理解任务，尤其是问答任务需要处理的问题；将语义角色标签用于机器阅读理解任务完全合理。

2004年，Narayanan 和 Harabagiu [30] 提出在问答（QA）系统中使用语义角色；这是最早将语义角色标签作为问答辅助任务的研究之一。此后，大多数将语义角色作为问答流程主要方法的研究 [31], [32]都取得了诸多积极成果。随着预训练语言模型 [9]–[12],占据主导地位，语义角色标签所提供的语义信息已不再受到重视或得到利用。然而，语义标签在许多其他自然语言处理应用中仍然很有帮助 [33], [34]。

2020年，Zhang 等人 [14] 提出，融入来自预训练语义角色标注任务的显式上下文语义，并引入一种更有效的语言表示模型，名为 SemBERT。该模型在英语的十项阅读理解和语言推理任务上取得了最先进的结果，并显著提升了表现。

目前尚无研究利用语义标签中的语义信息来增强越南语机器阅读理解/问答任务或自然语言理解任务。为了比较语义对越南语机器阅读理解任务中可回答性分类的影响，我们将采用 SemBERT 的方法，基于若干预训练语言模型构建不同版本的 SemBERT。

## III. 方法论

### A. *Retro-Reader*模型

回顾式阅读器（Retro-Reader） [8] 有两个并行模块，可用于执行两阶段阅读过程：一个是粗略阅读模块，另一个是精读模块。随后，将可回答性置信分数（精读模块）与判断分数（粗略阅读模块）相结合，得出最终答案，称为后验验证。在这两个模块中， [8]均采用基于 BERT 的预训练语言模型作为编码器，以获取输入词元的上下文表示。Retro-Reader模型的概览如图1所示。

粗略阅读模块可视为用于验证问题可回答性的分类器；最终隐藏状态经全连接层处理后，得到由可回答和不可回答元素组成的分类逻辑值。输出分数由无答案分数和答案分数组成。

精读模块旨在验证是否可回答、预测答案跨度，并给出最终预测答案。使用最后一层隐藏状态预测每个词元的开始概率和结束概率，从而得出预测的答案跨度和无答案分数。此外，还设有一个内部前置验证器，使精读模块也能够识别不可回答问题。

后验验证结合上述两个模块的输出，对最终答案进行聚合验证。

### B. 语义感知 *BERT*

语义感知 BERT (SemBERT) [14] 模型有三个主要组件，如图 3 所示：1）用于为输入句子标注各种语义角色标签的语义角色标注器；2）序列编码器，它使用预训练的语言表示模型生成

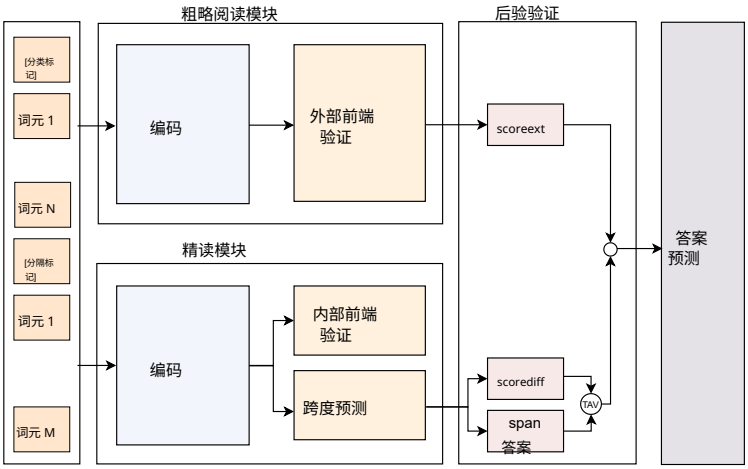


图 1. 回顾式阅读器模型概览 [8]

输入原始文本的上下文表示，并并行映射语义标签以创建语义表示；3）语义集成组件，它将文本表示与显式的上下文语义嵌入相结合，为下游任务生成集成表示。

语义角色标注是指使用预训练语义标注器，为具有不同语义序列的输入句子分配语义角色的过程。对于特定句子，会存在许多谓词-论元结构，例如图3所示的结构。

编码是通过 BERT 模型学习上下文表示和通过输入的语义角色标签学习语义表示的任务，包括两个并行过程。

- 上下文嵌入：BERT 模型在子词级别学习输入文本的上下文表示，然后逐词组合子词表示，并使用带最大池化的卷积神经网络（CNN）获取词级别的上下文表示。

- 语义嵌入：将语义角色标注器获得的这些标签序列通过查找表映射为向量，并输入双向门控循环单元层，以获取潜在空间中的表示；随后将其连接并输入全连接层，以获得语义表示。

集成：将词级别的上下文表示与基于语义角色标签的语义表示连接起来，形成下游任务的共同表示。

## IV. 实验与结果

### A. 数据集

*1)* 语义角色标注任务：我们使用 LORELEI 越南语语言包 [35]的语料库开展语义角色标注任务。该数据收集自讨论论坛、留言板、文档、社交网络和在线数据，包括单语和双语文本、词汇和标注。该数据集包含 1760 个使用简单语义标注 [36]进行标注的句子。

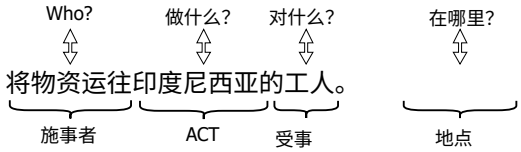


图 2. 带有 SSA 标注的越南语语义角色标注示例

### B. 语义感知 *BERT*

*2)* 机器阅读理解任务：UIT-ViQuAD 是越南语问答数据集，专为基于从越南语维基百科文章中提取的段落进行越南语机器阅读理解任务而创建。最初，UIT-ViQuAD1.0 [20]仅包含可回答的问题。为扩大机器学习规模，UIT-ViQuAD2.0[26]在 UIT-ViQuAD1.0 的 23K 个问题基础上，加入了 12K 多个无法回答的问题。

我们将机器阅读理解任务的数据集划分为训练集、开发集和测试集；各数据集的数量详见表I。

表 I 机器阅读理解数据集的概览统计。

| | 训练集 | Dev | Test | All |
|---|---|---|---|---|
| 文章数量 | 111 | 14 | 13 | 138 |
| 段落数量 | 3,038 | 637 | 426 | 4,101 |
| 问题数量 | 21,234 | 3,100 | 4,123 | 28,457 |
| 不可回答问题数量问题 | 6,890 | 1,359 | 968 | 9,217 |

### C. 评估指标

为了评估模型在越南语机器阅读理解任务中的表现，我们采用两项指标。

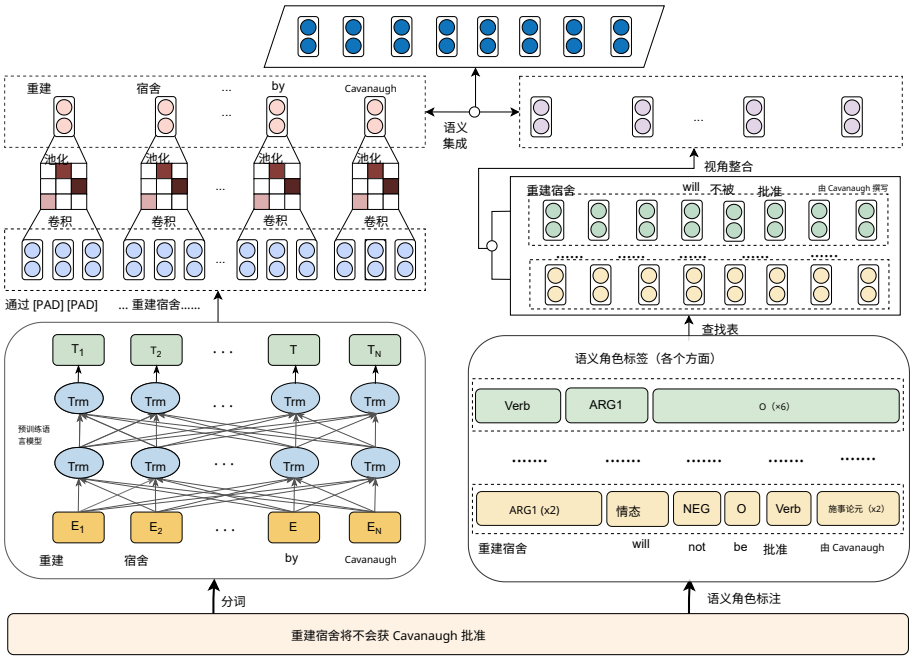


图3. 语义感知 BERT 模型概览 [14]

精确匹配（***EM***）是精确答案的数量：当预测与正确答案相同时得 1 分，否则得 0 分。在评估否定问题时，如果系统预测任何文本片段作为答案，则该问题自动得 0 分。

***F1*** 分数 通过比较预测答案中的各个词元与标准答案中的词元来估算，其依据是预期答案与标准答案之间匹配的词元数量。

$$Precision = \frac{N_{matched}}{N_{predicted}}$$

$$Recall = \frac{N_{matched}}{N_{gold_standard}}$$

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall}$$

其中，$N_{matched}$ 表示匹配的词元数量。$N_{predicted}$和 $N_{gold_standard}$ 分别表示预测答案中的词元数量和标准答案中的词元数量。

### *D.* 微调语义角色标注模型

在角色标注任务中，一个句子可能包含多个谓词-论元。因此，我们采用 K 折交叉验证方法，使用 $k = 10$ 训练了十个独立的语义角色标注模型，以从句子中提取语义信息的各个方面。对于每一折，我们对 XLM-RoBERTa 模型进行了 40 个 epoch 的微调，并使用 AdamW [37] 进行优化，学习率为 $1e-5$，权重衰减为 0.1，批次大小为 4。

### *E. Retro Reader*

我们使用了单语和多语预训练语言模型，例如基础版本的 mBERT、XLM-RoBERTa 和 PhoBERT。

- 对于粗略阅读模块，我们使用以下超参数：学习率为 $5e-6$，采用 AdamW 优化器，批次大小为 64，累积梯度为 8。
- 对于精读模块，我们采用 AdamW 优化器，学习率为 $2e-5$，批次大小为 64。输入词元的最大长度 max_seq_length 为语言表示模型的最大长度，而查询的最大长度 max_query_length 为 64。

最后，为了整合上述两个模块，我们还尝试调整上述两个模型之间的比率参数；阈值则根据开发集选定。

此外，我们使用 SemBERT 作为语言表示模型，并将其及其变体整合到 Retro-sketchy Reader 的阅读模块中，使用 SemBERT 的最终输出表示进行微调。我们采用 Retro-Reader模型的超参数。

## V. 结果与讨论

表II展示了各模型在 UIT-ViQuADv2.0 数据集上的性能评估结果。因此，我们可以清楚地看到，对于 Retro-Reader模型，使用语义与不使用语义相比，效果存在明显差异。使用语义使模型整体评估得分提高了 0.5–1%，体现在 F1_分数和精确匹配指标上。这一提升在统计上具有显著性，$p < 0.01$ 采用配对 t 检验。

各模型的粗略阅读模块在可回答性问题分类任务中的结果如表所示

表 II 机器阅读理解模型的结果表。符号‡表示与 Retro-Reader 使用配对 t 检验得到的结果相比，改进在统计学上显著，P < 0.01。

| 模型 | | 精确匹配 (%) | F1 分数（%） |
|---|---|---|---|
| 机器阅读理解基线 | mBERT | 42.130 | 55.000 |
| | XLM-R | 45.234 | 57.872 |
| | PhoBERT | 46.859 | 60.191 |
| Retro-Reader | mBERT | 44.118 | 55.858 |
| | XLM-R | 48.072 | 60.521 |
| | PhoBERT | 48.363 | 61.262 |
| Retro-Reader +语义（分类标记） | mBERT | 44.967‡ | 56.326‡ |
| | XLM-R | 48.485‡ | 61.065‡ |
| | PhoBERT | **49.032**‡ | **61.896**‡ |

III。正如我们所见，与传统上下文模型相比，在任务中为 mBERT、XLM-RoBERTa 和 PhoBERT 这三种语言模型应用语义角色标签中的语义后，它们识别不可回答问题的效果分别提升了 4.341%、1.9% 和 1.414%。这有助于提高 Retro-Reader模型在机器阅读理解任务中的总得分。

表 III 粗略阅读模块的不可回答问题验证任务结果。

| | 准确率（%） | F1 分数（%） |
|---|---|---|
| BERT | 71.258 | 48.275 |
| BERT +语义 | 73.490 | 55.518 |
| XLMR | 76.110 | 62.188 |
| XLMR +语义 | 76.985 | 66.695 |
| PhoBERT | 73.805 | 53.333 |
| PhoBERT +语义 | 75.036 | 56.954 |

根据表 IV 中语义角色标注任务的结果，我们可以看到各模型的 F1 分数较低。这可能是因为我们使用的 LORELEI 数据集中的语义角色标注数据量相对有限，不足以训练出优秀的标注模型。这会影响对输入句子的标注，而不准确的语义标签会损害语义信息的提取；因此，我们预计，如果训练出更准确的语义角色标注模型，这些改进效果可能会显著提升。

总体而言，Retro-Reader模型的提升幅度为 0.5% 至 1%，对于越南语机器阅读理解任务而言并不显著，但我们的实验在一定程度上展现了从越南语语义角色标注中提取语义的潜力。本实验的积极结果将为未来研究在越南语的其他自然语言理解任务中使用语义角色标注奠定基础。

## VI. 结论

### *A.* 总结

在本研究中，我们主要比较将语义信息融入 Retro-Reader 粗略阅读模块的编码器后，对越南语机器阅读理解问题中可回答性问题分类的影响。结果表明，加入语义信息的 mBERT、XLM-R 和 PhoBERT 基础版本模型，表现优于传统的上下文表示模型。我们介绍了该机器阅读理解问题的实现方法和实验结果。具体而言，在对问题可回答性进行分类这一子任务中，Retro-Reader模型的粗略阅读模块性能提升了 1.4 - 4.3%，使模型整体结果提升了 0.5 - 1%。这是越南语领域较早利用语义标签中的语义信息进行语言表示的研究之一。根据获得的实验结果，可以看出，从语义角色标注中提取语义信息，有望应用于越南语的各类自然语言理解任务。

表 IV 使用 LORELEI 数据集对 XLM-R 进行微调后，语义角色标注问题的结果。

| | 精确率 (%) | 召回率 (%) | F1 分数 (%) |
|---|---|---|---|
| 第 1 折 | 35.261 | 32.568 | 33.871 |
| 第 2 折 | 34.700 | 30.375 | 32.393 |
| 第 3 折 | 38.776 | 32.095 | 35.120 |
| 第 4 折 | 41.958 | 29.557 | 34.682 |
| 第 5 折 | 38.821 | 31.728 | 34.917 |
| 第 6 折 | 40.144 | 34.901 | 34.900 |
| 第 7 折 | 36.719 | 32.192 | 34.307 |
| 第 8 折 | 39.535 | 24.286 | 36.724 |
| 第 9 折 | 39.241 | 32.518 | 35.564 |
| 第 10 折 | 39.448 | 34.014 | 36.530 |
| 平均值 | 38.460 | 31.423 | 34.901 |

然而，在本实验的实施过程中，由于语料库规模有限，我们在训练预训练语义角色标注模型时遇到了一些困难。这影响了输入句子的语义角色标签标注。鉴于本研究取得的鼓舞人心的结果，我们希望能在一定程度上推动越南语语义角色标注任务的进一步研究，因为目前针对越南语这项任务的研究和语料库仍然很少。

### *B.* 未来工作

根据研究结果和局限性，我们确定了若干进一步研究和发展的方向，包括：

- 由于时间和资源有限，我们仅测试了预训练语言模型的基础版本和少量变体。未来，我们将尝试更多规模更大的语义感知 BERT 模型变体。
- 在本研究中，我们仅着重于改进可回答性分类任务；未来，我们将测试语义集成到 Retro-Reader 的精读模块中的效果。
- 我们认为，如果有足够的数据来训练 SRL 模型，输入句子标注的效果就会提升，从而使语义集成模型表现得更好。因此，未来我们将创建一个语义

角色标注数据集，用于训练比现有模型更准确的 SRL 模型。

## 致谢

本研究由胡志明市越南国家大学信息技术大学资助，项目编号为D1-2022-47。

## 参考文献

[1] K. M. Hermann、T. Kocisky、E. Grefenstette、L. Espeholt、W. Kay、M. Suleyman 和 P. Blunsom，“教机器阅读和理解”，神经信息处理系统进展，第28卷，2015年。[2] M. Joshi、E. Choi、D. S. Weld 和 L. Zettlemoyer，“TriviaQA：用于阅读理解的大规模远程监督挑战数据集”，arXiv预印本 arXiv:1705.03551，2017年。[3] M. Lewis 和 A. Fan，“生成式问答：学习回答完整问题”，载于国际学习表征会议，2018年。[4] J. Lee、S. Yun、H. Kim、M. Ko 和 J. Kang，“通过对段落进行排序改进开放域问答中的答案召回率”，arXiv预印本 arXiv:1810.00494，2018年。[5] P. Rajpurkar、R. Jia 和 P. Liang，“了解你不知道的内容：SQuAD中的不可回答问题”，arXiv预印本 arXiv:1806 03822，2018年。.[6] M. Hu、F. Wei、Y. Peng、Z. Huang、N. Yang 和 D. Li，“Read+Verify：带有不可回答问题的机器阅读理解”，载于AAAI人工智能会议论文集，第33卷，第01期，2019年，第6529–6537页。[7] S. Back、S. C. Chinthakindi、A. Kedia、H. Lee 和 J. Choo，“NeurQuri：用于机器阅读理解可回答性预测的神经问题需求检测器”，载于国际学习表征会议，2019年。[8] Z. Zhang、J. Yang 和 H. Zhao，“用于机器阅读理解的回顾式阅读器”，载于AAAI人工智能会议论文集，第35卷，第16期，2021年，第14 506–14 514页。[9] M. E. Peters、M. Neumann、M. Iyyer、M. Gardner、C. Clark、K. Lee 和 L. Zettlemoyer，“深度上下文化词表示”，载于《2018年计算语言学协会北美分会会议论文集：人类语言技术》，第1卷（长篇论文）。新奥尔良，路易斯安那州：计算语言学协会，2018年6月，第2227–2237页。

[10] [ ] 18 1202 在线提供：https://aclanthology.org/N - A. Radford、K. Narasimhan、T. Salimans、I. Sutskever 等，“通过生成式预训练改进语言理解”，2018年。[11] J. Devlin、M.-W. Chang、K. Lee 和 K. Toutanova，“BERT：用于语言理解的深度双向变换器预训练”，arXiv预印本 arXiv:1810.04805，2018年。[12] Y. Liu、M. Ott、N. Goyal、J. Du、M. Joshi、D. Chen、O. Levy、M. Lewis、L. Zettlemoyer 和 V. Stoyanov，“RoBERTa：一种稳健优化的BERT预训练方法”，arXiv预印本 arXiv:1907.11692，2019年。[13] Z. Lan、M. Chen、S. Goodman、K. Gimpel、P. Sharma 和 R. Soricut，“ALBERT：用于语言表示自监督学习的轻量级BERT”，arXiv预印本 arXiv:1909.11942，2019年。[14] Z. Zhang、Y. Wu、H. Zhao、Z. Li、S. Zhang、X. Zhou 和 X. Zhou，“语义感知BERT用于语言理解”，载于AAAI人工智能会议论文集，第34卷，第05期，2020年，第9628–9635页。[15] P. T. Hoang 和 P. X. Khoai，“构建越南语语义角色标注工具包”，2015年。[16] T.-H. Pham、X.-K. Pham 和 P. Le-Hong，“构建越南语语义角色标注系统”，载于2015年第十届数字信息管理国际会议（ICDIM）。电气电子工程师学会，2015年，第77–84页。[17] H. P. Le、T. H. Pham、X. K. Pham、T. M. H. Nguyen、T. L. Nguyen 和 M. H. Nguyen，“越南语语义角色标注”，2018年。[18] W. Lehnert，“问答过程。研究报告第88号。”1977年。[19] P. Rajpurkar、J. Zhang、K. Lopyrev 和 P. Liang，“SQuAD：100,000+个用于文本机器理解的问题”，arXiv预印本 arXiv:1606.05250，2016年。

[20] K. Nguyen、V. Nguyen、A. Nguyen 和 N. Nguyen，“用于评估机器阅读理解的越南语数据集”，载于第28届国际计算语言学会议论文集，2020年，第2595–2605页。[21] K. Van Nguyen、K. V. Tran、S. T. Luu、A. G.-T. Nguyen 和 N. L.-T. Nguyen，“利用外部知识增强基于词汇的方法，以实现越南语多项选择机器阅读理解”，IEEE Access，第8卷，第201 404–201 417页，2020年。[22] K. Van Nguyen、T. Van Huynh、D.-V. Nguyen、A. G.-T. Nguyen 和 N. L.- T. Nguyen，“用于健康新闻文章机器阅读理解的新越南语语料库”，《亚洲及低资源语言信息处理汇刊》，2020年。[23] R. Kadlec、M. Schmid、O. Bajgar 和 J. Kleindienst，“使用注意力求和阅读器网络进行文本理解”，arXiv预印本 arXiv:1603 01547，2016 年。.[24] W. Wang 、 N. Yang 、 F. Wei 、 B. Chang 和 M. Zhou，“用于阅读理解和问答的门控自匹配网络”，载于第55届计算语言学协会年会论文集（第1卷：长篇论文），2017年，第189–198页。[25] A. Trischler、T. Wang、X. Yuan、J. Harris、A. Sordoni、P. Bachman 和 K. Suleman，“NewsQA：机器理解数据集”，arXiv预印本 arXiv:1611.09830，2016年。[26] N. V. Kiet、T. Q. Son、N. T. Luan、H. V. Tin、L. T. Son 和 N. L.-T. Ngan，“VLSP 2021 - ViMRC 挑战赛：越南语机器阅读理解”，《越南国家大学科学杂志：计算机科学与通信工程》，第38卷，第1期，2022年。[27] X. Liu、W. Li、Y. Fang、A. Kim、K. Duh 和 J. Gao，“用于SQuAD 2.0的随机答案网络”，arXiv预印本 arXiv:1809.09194，2018年。[28] Z. Li、S. He、H. Zhao、Y. Zhang、Z. Zhang、X. Zhou 和 X. Zhou，“依存关系还是跨度：端到端统一语义角色标注”，载于AAAI人工智能会议论文集，第33卷，第01期，2019年，第6730–6737页。[29] Z. Li、S. He、J. Cai、Z. Zhang、H. Zhao、G. Liu、L. Li 和 L. Si，“用于语义角色标注的统一语法感知框架”，载于2018年自然语言处理实证方法会议论文集，2018年，第2401–2411页。[30] S. Narayanan 和 S. Harabagiu，“基于语义结构的问答”，国际计算机科学研究所，伯克利，加利福尼亚州，技术报告，2004年。[31] G. Fliedner，《语言学驱动的问答》。德国人工智能研究中心，萨尔大学，2007年。[32] B. Ofoghi、J. Yearwood 和 R. Ghosh，“利用词汇资源中的封装语义构建混合问答架构”，载于澳大拉西亚人工智能联合会议。施普林格，2006年，第1276–1280页。[33] T. Mihaylov 和 A. Frank，“基于词嵌入的跨论元语义相似度进行话语关系类别分类”，载于CoNLL-16共享任务论文集，2016年，第100–107页。[34] C. Shi、S. Liu、S. Ren、S. Feng、M. Li、M. Zhou、X. Sun 和 H. Wang，“用于机器翻译的基于知识的语义嵌入”，载于第54届计算语言学协会年会论文集（第1卷：长篇论文），2016年，第2245–2254页。[35] J. Tracey、S. Strassel、D. Graff、J. Wright、S. Chen、N. Ryant、S. Kulick、K. Griffitt、D. Delgado 和 M. Arrigo，“LORELEI越南语代表性语言包”，2022年。 [在线]提供：https://hdl.handle.net/11272.1/AB2/JWPEIA [36] K. Griffitt、J. Tracey、A. Bies 和 S. Strassel，“简单语义标注与情境框架：LORELEI中基础文本理解的两种方法”，载于LREC，2018年。[37] I. Loshchilov 和 F. Hutter，“解耦权重衰减正则化”，载于第7届国际学习表征会议，ICLR 2019，新奥尔良，路易斯安那州，美国，2019年5月6日至9日。OpenReview.net，2019年。[在线]提供：https://openreview.net/forum?id=Bkg6RiCqY7