Computer Science and Application 计算机科学与应用, 2026, 16(1), 115-129
Published Online January 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.161010

Hans汉斯

基于图神经网络和知识图谱的可解释小样本文本分类模型

周子璇，李秋瑶

河北地质大学信息工程学院，河北 石家庄

收稿日期：2025年12月5日；录用日期：2026年1月5日；发布日期：2026年1月14日

摘　要

小样本文本分类具有广泛的应用场景。然而，现有方法面临两个关键挑战：数据稀缺和可解释性不足。为此，本实验提出ARExplainer方法，这是一个数据与推理增强的可解释小样本文本分类方法。通过利用大语言模型(LLMs)的泛化能力，有效扩展了训练样本的多样性，从而缓解了小样本学习的数据瓶颈。针对模型可解释性问题，构建了知识图谱驱动的推理引擎。该引擎结合图注意力网络提取可验证的符号推理路径，为分类决策提供逻辑依据，最后，利用基于提示的解释生成器生成简洁、清晰的自然语言解释。实验结果表明，在1-shot设置下，ARExplainer显著优于最好的基线模型。此外，通过与自动生成解释和人工标注结果的对比分析，证实ARExplainer能够提供更便于人类理解的自然语言解释。

关键词

小样本学习，大语言模型，知识图谱，可解释文本分类模型，数据增强

An Interpretable Few-Shot Text Classification Model Based on Graph Neural Networks and Knowledge Graphs

Zixuan Zhou, Qiuyao Li

School of Information, University of Hebei GEO University of China, Shijiazhuang Hebei

Received: December 5, 2025; accepted: January 5, 2026; published: January 14, 2026

文章引用: 周子璇, 李秋瑶. 基于图神经网络和知识图谱的可解释小样本文本分类模型[J]. 计算机科学与应用, 2026, 16(1): 115-129. DOI: 10.12677/csa.2026.161010

Abstract

Few-shot text classification has broad application scenarios. However, existing methods face two key challenges: data scarcity and insufficient interpretability. To address these issues, this paper proposed ARExplainer, an interpretable few-shot text classification method with data and reasoning augmentation. By leveraging the generalization capability of Large Language Models (LLMs), it effectively expanded the diversity of training samples, thereby alleviating the data bottleneck in few-shot learning. For the model interpretability issue, the method constructed a knowledge graph-driven reasoning engine. This engine combined a Graph Attention Network to extract verifiable symbolic reasoning paths, providing logical evidence for classification decisions. Finally, it utilized a prompt-based explanation generator to produce concise and clear natural language explanations. Experimental results demonstrated that ARExplainer significantly outperformed the strongest baseline model in the 1-shot setting. Furthermore, comparative analysis against automatically generated explanations and human-annotated results confirmed that ARExplainer provides natural language explanations that are easier for humans to understand.

Keywords

Few-Shot Learning, Large Language Models, Knowledge Graph, Interpretable Text Classification Model, Data Augmentation

1. 引言

在自然语言处理(NLP)领域，文本分类广泛应用于情感分析、信息检索和问答系统等任务。然而，随着任务复杂性的增加，小样本学习中的数据稀缺问题成为瓶颈，传统方法依赖大量标注数据，但获取这些数据昂贵且困难。现有的主要解决方法包括元学习、迁移学习和数据增强，它们分别通过任务间的知识共享、预训练模型的微调以及样本生成来缓解数据不足问题[1]。元学习方法通过任务间的知识共享实现快速适应，但在跨领域迁移时性能显著下降；迁移学习方法(如 BERT 微调)，依赖预训练模型的表征能力，但在样本量少于 10 个/类时容易过拟合，数据增强方法(如 EDA)，通过文本改写扩充样本，但生成的样本多样性有限，词汇重叠率极高。近年来，基于提示调优(Prompt Tuning)的方法成为主流[2]，它通过将下游任务与预训练语言模型的任务对齐，充分利用预训练模型的知识，避免大量额外参数训练。然而，这一方法仍面临可解释性不足的问题，尤其在特定领域中，如金融、法律、医疗，模型的决策过程难以理解，限制了其广泛应用。

目前，大型语言模型(LLMs)如 BERT [3]、RoBERTa [4]和 T5 [5]在预训练后，已在各种自然语言处理任务中表现出色。然而，随着模型规模的急剧增加，虽然这些模型在处理复杂任务时展示了巨大潜力[6]，但它们仍面临显著的可解释性问题。LLMs 通过记忆训练数据中的知识，但缺乏有效的事实回忆，经常生成虚假或不准确的信息[7]-[9]。此外，LLMs 的推理过程是基于概率模型，缺乏透明性[10]，且其决策模式无法直接解释给人类[11]。为解决这些问题，研究者提出了多种解释方法，主要包括内在方法和事后方法[12] [13]。内在方法依赖于简单模型，如线性模型；事后方法则在模型训练后提供特征选择等解释。尽管一些方法如思维链和反事实解释[14]尝试解释推理过程，但这些方法往往未能提供对推理过程的全面

理解。因此，要实现模型的高效解释，不仅需要深入挖掘推理的原因，还应以可读的格式呈现文本解释，从而帮助人类全面理解和信任模型的决策过程。

为了解决 LLMs 的可解释性和小样本学习问题，本研究提出了 ARExplainer 方法，结合了 LLMs 和知识图谱(KGs)的优势，实现了数据与推理增强。本研究首先利用提示工程设计数据增强机制，扩展训练样本，增强模型的生成能力实现了数据的增强；其次，构建知识图谱推理引擎，由于知识图谱是通过三元组存储结构化知识，所以可以提供准确且可解释的显式信息，并具有符号推理能力。所以我们结合知识图谱和增强后的数据，利用图神经网络提取多跳推理路径，实现了模型推理的增强，使模型能够基于检索路径进行忠实推理并生成自然语言的可解释的结果。

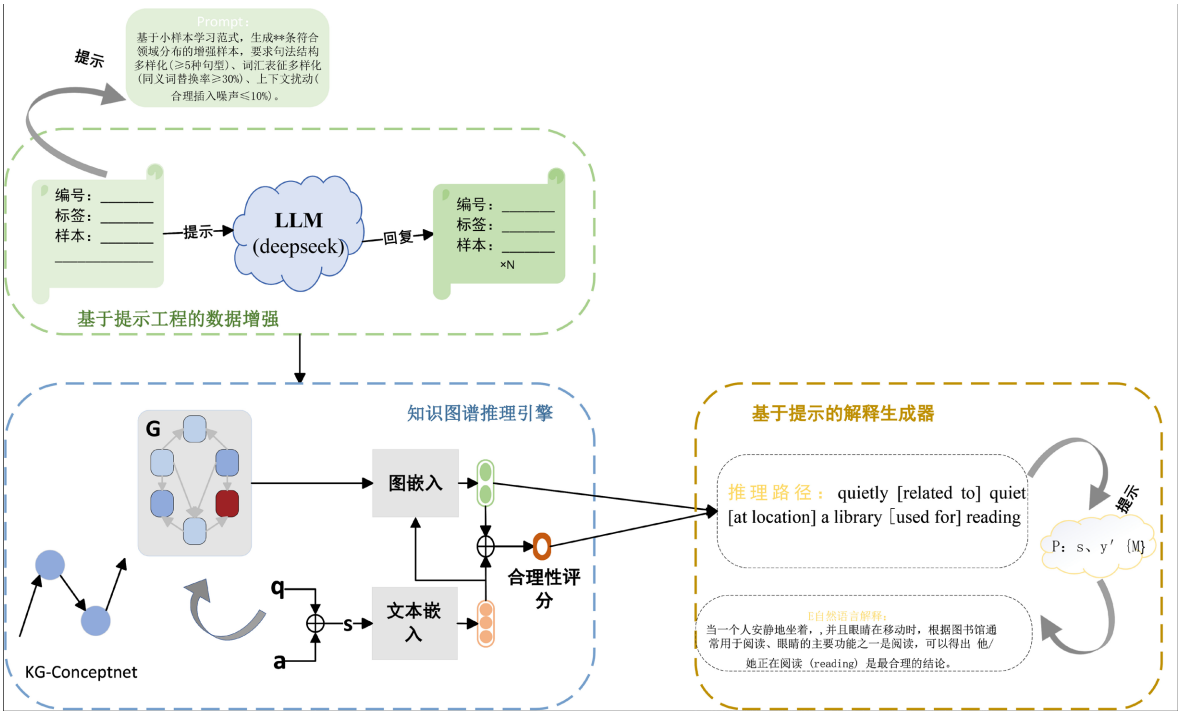

Figure 1. ARExplainer model architecture diagram
图 1. ARExplainer 模型架构图

2. 相关工作

2.1. 基于大模型的数据增强

数据增强(DA)通过生成新训练样本扩展数据集，提升模型的泛化能力，尤其在小样本学习和自然语言处理(NLP)任务中尤为关键。然而，高质量数据的获取面临成本高、耗时长、标注繁琐且易出错[15]等挑战。传统数据增强方法(如同义词替换、回译)受限于数据质量和多样性，而大型语言模型(LLMs)能够生成高质量文本，有效弥补这些不足[16]。研究表明，LLM 生成的增强数据在情感分析和仇恨言论检测等任务中显著提升了分类性能[17] [18]。然而，LLMs 生成的数据可能存在重复性和低多样性的问题，影响训练效果。针对这一点，Liu [19]等人提出可控数据增强，通过定制化提示提升数据多样性。

从数据角度来看，LLMs 为克服数据稀缺提供了可行策略，可生成高质量合成数据，甚至在某些情况下优于人工标注数据。这一方法不仅缓解了标注数据短缺问题，还能在计算成本较低的情况下扩展训练集规模[20]。合成数据的应用降低了数据收集成本和能耗，推动了模型训练和推理的发展。总体而言，基

于 LLMs 的数据增强是小样本学习的有效解决方案，但提升生成数据的质量和多样性仍是未来研究的重点。

2.2. 基于提示调优的文本分类

在自然语言处理领域，基于提示学习[21] (Prompt-based Learning)作为一种新兴的学习范式，逐渐发展并展现出独特优势。与传统监督学习方法不同，提示学习[22]依赖于语言模型，通过直接建模文本的概率，生成填充槽位的文本提示，从中推导出最终输出。该方法支持大规模预训练，并能够通过定义新的提示函数进行少量样本或零样本学习，快速适应新场景，且几乎不依赖于标签数据。提示工程的核心目标是设计最有效的提示函以提高任务表现，通常需要根据任务需求选择合适的提示形态。提示模板可以是手动创建[23]，也可以通过自动化模板学习生成，后者通过自动搜索优化提示模板，减少人工设计的局限，尤其在复杂任务中表现出色。自动生成的提示包括离散提示[24] [25]自然语言和连续提示[26] [27]在嵌入空间中执行，使模型能够在数据稀缺的环境下高效优化任务表现。

2.3. 知识图谱增强大模型

近年来，LLMs 与知识图谱(KGs)结合的研究引起了广泛关注。LLMs 和 KGs 本质上是相辅相成的，可以互相增强。在 KG 增强型 LLMs 中，知识图谱不仅可以在 LLMs 的预训练和推理阶段提供外部知识[28]-[30]，还能提升 LLMs 的可解释性[31] [32]。虽然 LLMs 通过从大规模语料库中学习知识并在各种 NLP 任务中取得先进性能，但常常受到幻觉问题和缺乏可解释性的批评。为了应对这些问题，研究者将 KGs 引入 LLMs，以增强其知识感知能力，并提高推理的准确性。KGs 可以在预训练阶段帮助 LLMs 学习知识[33] [34]，或在推理阶段通过检索知识增强领域特定知识的获取。此外，KGs 还可以用来解释 LLMs 的事实和推理过程。尽管将 KG 融入 LLM 预训练有效地融合了知识与文本表示，但这些方法无法动态更新知识，且在推理过程中难以处理未见过的新知识[35]。因此，许多研究集中于在推理阶段注入知识，并保持知识空间与文本空间的分离，尤其是在问答任务中，这对于捕捉文本语义和最新现实知识至关重要。

3. 提出的方法：AExplainer

本文模型的整体架构如图 1 所示，主要包含以下三个核心步骤：

a) 基于提示工程的数据增强机制(第 3.2 节)：针对小样本数据，利用提示在大语言模型上进行数据增强，以提高数据的多样性和模型的泛化能力。

b) 知识图谱推理引擎(第 3.3 节)：通过引入知识图谱(KG)并采用图神经网络(GNN)，提升模型的推理能力。此外，结合显式推理方法，识别和提取影响模型预测的关键推理因素，从而增强推理的可解释性。

c) 基于提示的解释生成(第 3.4 节)：采用基于提示生成方法，利用前一步识别的推理元素，生成清晰的文本解释，展现模型的决策过程。

这一架构通过数据增强提升模型的泛化能力，融合知识图谱增强推理，并结合可解释性机制，使模型的推理过程更加透明和可追溯。

3.1. 任务定义

研究旨在解决小样本文本分类任务中数据稀缺与模型可解释性的双重挑战：

a) 如何在小样本约束下实现数据增强提升分类性能？

b) 如何通过结构化知识增强的推理路径生成机制，为分类决策提供逻辑可验证的解释？

为此，形式化定义知识引导的可解释文本分类任务：给定输入文本序列 x 及其所属类别 $y \in C$，任务目标不仅需要准确预测 f_{LM}，还需生成结构化推理证据集合 $E = \left\{\left(\mathrm{e_i}, \mathrm{r_{ij}}, \mathrm{e_j}\right)\right\}$ 及其自然语言解释 E'。其中

任务可形式化为两阶段过程：

a) 增强分类：$H\left(x,LLM\right)\to X$；

b) 解释生成：$E'\leftarrow \text{Explanation}\left(f_{LM},s,a,E\right)$，通过图注意力机制提取关键推理路径并生成自然语言描述。

其中，x 是小样本，X 是增强后的样本，s 是上下文，a 是答案，解释 E' 是对预测背后的推理的洞察，并以人类可以理解的格式呈现。

3.2. 文本分类到问答任务的适配机制

为将知识图谱推理引擎应用于文本分类任务，本方法将分类问题重构为多项选择问答形式。给定输入文本 x 和候选类别集合 C，构造问题 q 为“该文本 x 属于以下哪个类别？”，每个选项 a_i 对应一个类别标签。然后将重构后的内容 $\left[q;C\right]$ 输入给模型。后续子图构建与推理过程均基于该陈述进行。该适配机制使模型能够利用知识图谱中的结构化知识进行类别推理，并为解释生成提供逻辑路径。

3.3. 基于提示工程的数据增强机制

正如引言中所述，由于 LLM 的规模，本方法在设计上注重兼容 LMaaS (语言模型即服务)，即仅通过 API 访问 LLM，无需微调模型或获取其嵌入向量及 logits。为实现这一目标，采用开放式查询策略，通过提示引导 LLM 进行数据增强，以充分挖掘其在文本形式中学习到的特征和通用知识，并将这些信息转化为下游 GNN 的有效节点特征。基于这一动机，本研究针对每个样本生成相应的提示，并根据具体任务和数据集定制提示内容。一般的提示格式如下图 2：

Figure 2. Prompt template
图 2. 提示模板

查询 LLM 会生成最后增强后的数据，如下图 3：

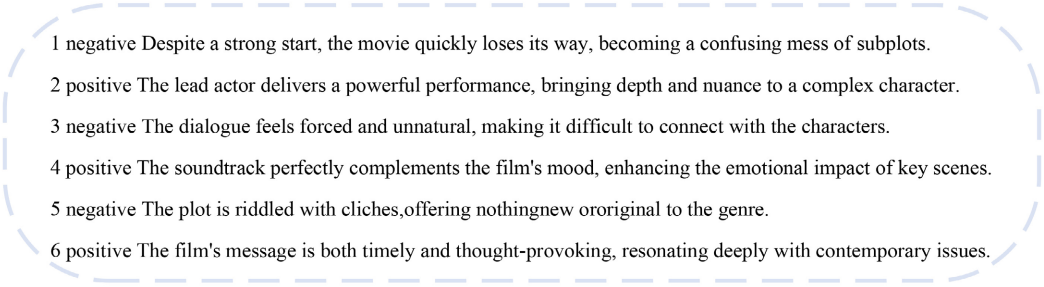

Figure 3. Form of augmented data
图 3. 增强后的数据形式

最后将这些增强后的数据，提供给下游的语言模型和图神经网络模型，具体内容将在下一节中描述。

3.4. 知识图谱推理引擎

3.4.1. 子图构建

本研究针对多项选择问答任务，利用外部知识图谱(KG)进行推理。核心目标是从给定选项中选出最合理的答案。将问题与每个选项连接，形成陈述并编码为表示 $s=[q;a]$。之后，我们使用 ConceptNet 作为通用领域的知识图谱，来测试模型利用结构化知识源的能力。首先合并 ConceptNet 关系类型以增加图的密度，并添加反向关系，构建了一个具有 34 种关系类型的多关系图。为了从知识图谱中提取一个信息丰富的上下文化图 G，我们在 s 中识别实体提及，并将其链接到 ConceptNet 中的实体，用其初始化我们的节点集 V。然后，我们将出现在任意提到的实体对之间的所有两跳路径中的实体添加到 V。我们不进行任何剪枝，而是保留节点 V 之间的所有边，形成我们的 G。

这个上下文化的子图被定义为一个多关系图 $(G=(V,E,\varnothing))$，其中 V 是外部 KG 中实体的子集，仅包含与 s 相关的实体。$E\subseteq V\times V$ 是连接 V 中节点的边的集合，$R=\{1,\cdots,m\}$ 是所有预定义关系类型的 ID 集合。映射函数 $\phi(\mathrm{i}):\mathrm{V}\to\mathrm{T}\{E_q,E_a,E_0\}$ 以节点 $i\in V$ 作为输入，输出 E_q 如果 i 是在 q 中提到的实体，输出 E_a 如果它在 a 中提到，否则输出 E_0。最后，将 G 编码为 g，并将 s 和 g 连接起来以计算合理性得分将陈述表示与子图信息结合，计算合理性得分，并选取得分最高的选项作为答案。这一方法有效利用外部知识增强推理能力，可推广至其他知识驱动任务。

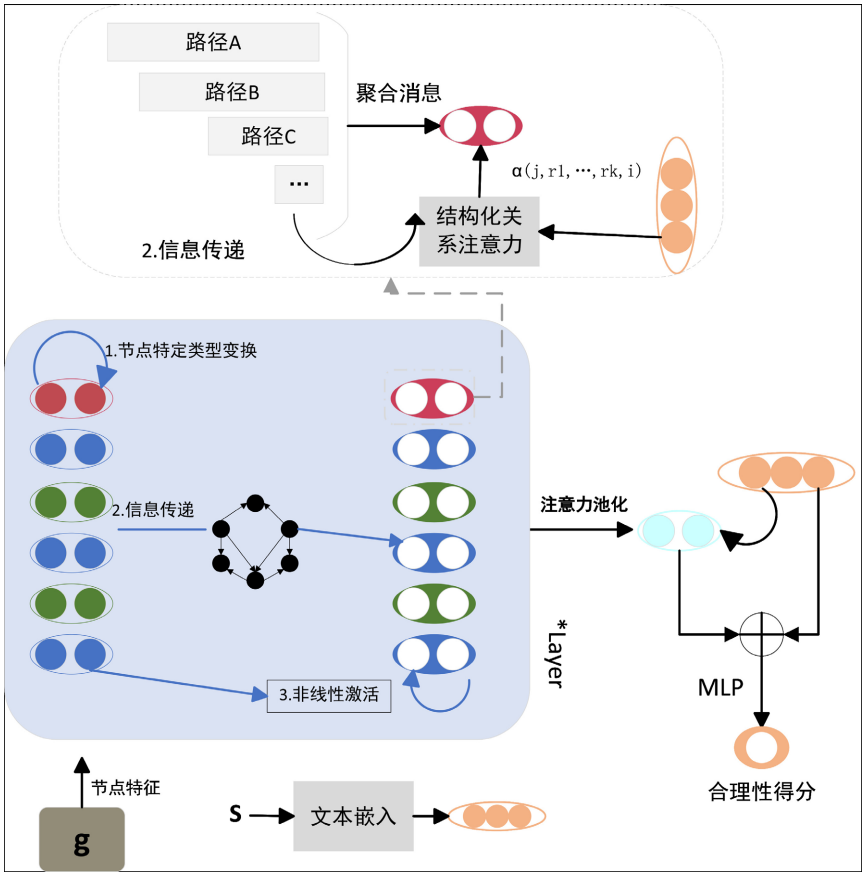

Figure 4. Reasoning engine architecture diagram
图 4. 推理引擎架构图

3.4.2. 知识图谱推理引擎架构

本研究将陈述 s 的编码留给预训练语言模型，推理部分遵循传统的 GNN 框架如图四，其中节点特征可以使用预训练权重进行初始化。这里重点讨论节点嵌入的计算，推理引擎架构图如图 4 所示。

特定类型的变换：为了使模型感知节点类型 φ，首先对输入节点特征执行节点类型特定的线性变换：

$$x_i = U_{\varphi(i)} h_i + b_{\varphi(i)} \tag{1}$$

其中，可学习的参数 U 和 b 是针对节点 i 的类型特定的。

多跳消息传递：如前所述，本研究动机是赋予 GNNs 直接建模路径的能力。为此，我们提出在所有长度不超过 K 的关系路径上直接传递消息。有效的 k 跳关系路径集定义为：

$$\Phi_k = \left\{ (j, r_1, \cdots, r_k, i) \middle| (j, r_1, j_1), \cdots, (j_{k-1}, r_k, i) \in \varepsilon \right\} (1 \le k \le K) \tag{2}$$

在这些路径上执行 k 跳 $(1 \le k \le K)$ 消息传递，这是对 RGCNs 中单跳消息传递(参见公式 3)的推广：

$$h_i' = \sigma\left(\left(\sum_{r \in R} \left| N_i^r \right| \right)^{-1} \sum_{r \in R} \sum_{j \in N_i^r} W_r h_j \right) \tag{3}$$

$$z_i^k = \sum \left(j, r_1, \cdots, r_k, i \right) \in \Phi_k \alpha \left(j, r_1, \cdots, r_k, i \right) \Big/ d_i^k \cdot W_0^K \cdots W_0^{k+1} W_{r_k}^k \cdots W_{r_1}^1 x_j \, (1 \le k \le K) \tag{4}$$

其中，矩阵 $W_r^k \, (1 \le k \le K, 0 \le r \le m)$ 是可学习的，$\alpha(j, r_1, \cdots, r_k, i)$ 是注意力得分，而 $d_i^k = \sum_{(j \to i) \in \varnothing_k} \alpha(j \to i)$ 是归一化因子。矩阵 $\left\{ W_{r_k}^k \cdots W_{r_1}^1 \middle| 1 \le r_1, \cdots, r_k \le m \right\}$ 可以解释为一个 $(m \times m)^k \times d \times d$ 张量的低秩近似，为每 k 跳关系分配一个单独的变换，其中 d 是 x_i 的维度。

来自不同长度路径的传入消息通过注意力机制进行聚合：

$$z_i = \sum_{k=1}^{K} soft\max \left(bilinear \left(s, z_i^k \right) \right) \cdot z_i^k \tag{5}$$

非线性激活。最后，应用捷径连接(shortcut connection)和非线性激活来获得输出节点的嵌入。

$$h_i' = \sigma \left(V h_i + V' z_i \right) \tag{6}$$

其中，V 和 V' 是可学习的模型参数，σ 是一个非线性激活函数。

在公式 7 中，直接对注意力得分 $\alpha(j, r_1, \cdots, r_k, i)$ 进行参数化，将会需要 $O(m^k)$ 个参数用于 k-跳路径。为了提高效率，首先将其视为关系序列的概率。

$$\alpha(j, r_1, \cdots, r_k, i) = p\left(\varphi(j), r_1, \cdots, r_k, \varphi(i) \middle| s \right) \tag{7}$$

这种关系序列的概率可以自然地通过概率图模型来建模，例如条件随机场：

$$\begin{aligned} p(\cdots | s) &\propto \exp\left(f(\varphi(j), s) + \sum_{t=1}^{k} \delta(r_t, s) + \sum_{t=1}^{k-1} \tau(r_t, r_{t+1}) + g(\varphi(i), s) \right) \\ &\triangleq \beta(r_1, \cdots, r_k, s) \cdot \gamma(\varphi(j), \varphi(i), s) \end{aligned} \tag{8}$$

其中，$\beta(r_1, \cdots, r_k, s)$ 是关系类型注意力，$\gamma((j), (i), s)$ 是节点类型注意力，$f(\cdot)$、$\delta(\cdot)$ 和 $g(\cdot)$ 由两层 MLP 进行参数化，而 $\tau(\cdot)$ 则由一个形状为 $m \times m$ 的转换矩阵进行参数化。直观上，$\beta(\cdot)$ 建模了一个 k-跳关系的重要性，而 $\gamma(\cdot)$ 则建模了从节点类型 $\varphi(j)$ 到 $\varphi(j)$ 的消息的重要性(例如，模型可以学习仅从问题实体向答案实体传递消息)。

模型通过将 k 跳关系分解为既考虑上下文的单跳关系(由 δ 建模)和两跳关系(由 τ 建模)来为其评分。

3.5. 基于提示的解释生成器

基于提示的解释生成分为两步如图 5，首先是对关键元素的提取，二是将提取到的关键元素作为提示集成到基于提示的解释生成器中。我们首先提取对模型决策过程至关重要的关键信息。这些关键元素包括最终答案、推理过程所对应的节点、边以及注意力权重(α)。本研究用 M 表示提取的关键元素。将输出表示为 E'。E' 是一个自然语言解释。

第二步将关键元素 $\{M\}$ 集成到基于提示的解释生成器中。基于提示的解释生成器依赖于一组预定义的结构来指导解释的生成。生成器包括输入样本 z、模型预测的输出 y' 和提取的关键元素 $\{M\}$，使用 DeepSeek 模型来提供模型推理过程的字面解释。生成器的输出是以句子或段落形式呈现的自然语言解释。

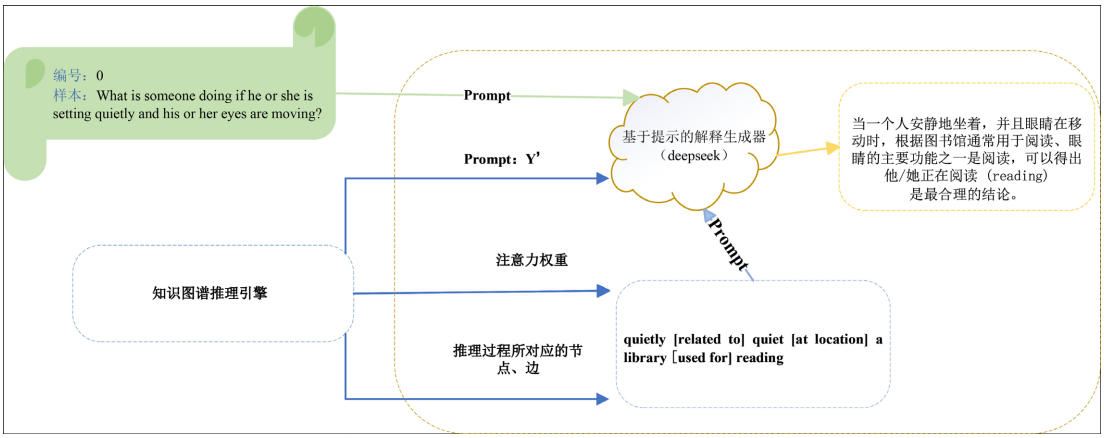

Figure 5. Case study of the prompt-based explanation generator
图 5. 基于提示的解释生成器实例说明

3.6. 学习和推理

现在本研究讨论的模型在问答任务中的学习和推理过程。根据上面的问题表述，目标是在给定问题 q 和文本 s 以及图 G 的信息的情况下，确定答案选项 $a \in C$ 的合理性。首先通过对答案实体的输出节点嵌入 $\{h_i | i \in A\}$ 进行注意力池化来获得图表示 g。接着，将其与文本表示 s 拼接，并通过 $\rho(q,a) = MLP(s \oplus g)$ 计算合理性得分。

在训练过程中，通过最小化交叉熵损失来最大化正确答案 $\hat{a}$ 的合理性得分：

$$L = E_{q,\hat{a},C}\left[-\log \frac{\exp(\rho(q,a))}{\sum_{a \in C}\exp(\rho(q,a))}\right] \tag{9}$$

整个模型与文本编码器(例如，RoBERTa)一起进行端到端的联合训练。

在推理过程中，通过 $\arg\max_{a \in C} \rho(q,a)$ 来预测最有可能的答案。此外，我们可以解码推理路径作为模型预测的证据。具体来说，就是确定在池化层中得分最高的答案实体 i^* 和公式 8 中得分最高的路径长度 k^*。然后，通过 $\arg\max \alpha(j, r_1, \cdots, r^*_k, i^*)$ 来解码推理路径，通过得到的这些关键信息，以及输入样本 z 和模型预测的输出 y' 作为提示，提示基于提示的解释生成器。

4. 实验

4.1. 数据介绍

本研究在四个文本分类数据集(SST-2、Amazon、AGNews、CommonsenseQA)上评估了模型性能，涵

盖情感分析、新闻分类和常识推理任务，全面考察了模型在小样本学习环境下的泛化能力。

a) **SST-2** 是经典的情感分析数据集，包含经过精细标注的电影评论句子，该任务要求模型判断文本的情感极性(正面或负面)，评估模型对短文本情感分类的表现。

b) **Amazon** 数据集则由电商评论构成，涵盖多个类别，其中不仅包含显式情感表达，还涉及隐含情绪、多义性词汇、讽刺表达等复杂语言现象，测试模型处理长文本和复杂语言现象的能力。

c) **AGNews** 由在线新闻数据构成，涉及 World、Sports、Business、Technology 科技四个类别。该数据集专注于新闻文本分类任务，要求模型能够准确识别新闻的主题，考察模型在不同文本领域中的主题识别和实体建模能力。

d) **CommonsenseQA** 专注于常识推理任务，旨在评估模型在多跳推理中的能力，我们仅用该数据集验证了知识图谱推理引擎对模型推理能力的提升，而不用于小样本性能验证。

4.2. 基线模型介绍

本研究的评估分为两个部分。在第一部分，关注模型性能。将 AExplainer 与基线模型在 SST-2、Amazon、AGNews 数据集上进行比较。基线模型包括如微调(FT)和提示调优(Prompt-tuning)，而动词化器(Verbalizers)是提示调优成功的关键，所以本实验在少样本设置下使用多种动词化器。这些方法包括离散型动词化器，旨在通过特定离散词语映射答案，如：手动动词化器(MV) [23]、自动动词化器(AV) [2]；连续型动词化器：在无限连续空间中搜索标签映射参数：如 WARP [36]和原型动词化器(PV) [37]。并使用 CommonsenseQA 数据集在仅预训练模型下的结果作为基线，来验证知识图谱推理引擎的推理能力。

在第二部分，评估 AExplainer 的解释能力。为了建立比较基准，本研究使用了两个先前的工作作为基准，即 PathReasoner 和 CommonsenseQA 的解释。这些工作因提供自然和易于理解的解释而被认可。

4.3. 实验设置和评价指标

本研究将 GNN 模块设置为 200 维和 5 层，每一层应用 0.2 的 dropout 率。在训练过程中，使用 RAdam 优化器，并在单个 NVIDIA RTX 4090 GPU 上进行模型训练。批量大小为 64，语言模型和 GNN 模块的学习率分别设置为 1e-5 和 1e-3。这些设置用于评估的第一部分，以研究 GNN 模块的性能。第二部分使用 deepseeek 模型来实现模型推理过程的字面解释。

使用 ConceptNet 作为 SST-2、Amazon、AGNews 和 CommonsenseQA 任务的外部知识源。在 1-shot 实验中，从训练集中随机选择每个类别的一个实例，作为提示，并将其传递给基于提示工程的数据增强机制，生成 1000 至 2000 个增强数据作为新的训练集。最终结果使用 micro-F1 作为评估指标。

4.4. 实验对比及分析

4.4.1. 提示对比及分析

为验证基于提示工程的数据增强机制中提示的作用，本实验将没有具体要求的提示的生成结果与之作对比。第一种提示(如图 6)生成速度较快，因为没有额外的复杂要求。然而，生成的文本缺乏足够的多样性，容易重复使用相似的句型和表达方式，词汇变化有限，导致数据表现出单一的表达方式，缺乏灵活性和变化，文本长度和复杂度较低，无法实现长文本的任务。与之相比，本实验使用的第二种提示(如图 7 所示)虽然生成速度较慢，但生成的文本在词汇和句法结构上表现出更高的多样性。通过同义词替换和句式变化，确保了数据集的丰富性。生成的文本长度超过 70 字，能够包含更多的细节和背景信息，使数据更符合实际应用场景，提升了其实际有效性。合理的噪声插入也增强了数据的适应性，使其更贴合现实中的不完美情况。

1 I would give this more stars if I could. good

2 This product arrived damaged and completely unusable. bad

3 The materials are eco-friendly and durable. good

4 Worst purchase I've made in years. Total waste of money. Bad

5 It's exactly what the photos showed. good

Figure 6. Prompt: “Please expand the data based on the text content and add 200 entries”
图 6. 提示词：“请你根据文本内容，进行扩充数据，补充 200 条数据”

1 This produet is absolutely fantastie! The quality is top-noteh and it exceeded all my expectations.Iwould highly recommend it to anyone looking for a reliable and durable item. Good

2 I was really disappointed with this purchase, The product broke after just a few uses and thecustomer service was unhelpful,I would not recommend this to anyone. Bad

3 The design of this item is sleek and modern, making it a great addition to my home.It functions perfecty and I couldn't be happier with my purchase. Good

4 This product is a complete waste ofmoney, It doesn't work as advertised and the materialsfeel cheap.I regret buying it. bad

5 I am thoroughly impressed with the performance ofthis product. It has made my dailytasks so much easier and more effcient, Definitely worth the investment. good

Figure 7. Prompt: “Require the text to be longer than 70 words, with diverse sentence structures (≥5 types of sentences), varied vocabulary (synonym replacement rate ≥30%), and contextual perturbation (reasonable noise insertion ≤ 10%).”
图 7. 提示词“要求字数多于 70 字，要求句法结构多样化(≥5 种句型)、词汇表征多样化(同义词替换率≥30%)、上下文扰动(合理插入噪声 ≤ 10%)”

4.4.2. 实验对比及分析

在表 1 中展示了实验结果，其中评估了本研究提出的模型方法在 SST-2、Amazon、AGNews 数据集上的准确性。实验结果表明，AExplainer 的与现有基线方法相比，在 SST-2、Amazon、AGNews 这三个数据集的小样本上的性能都有显著提升。因为模型只对小样本数据做数据增强，所以仅考虑了 1-shot 情况下各种方法的性能比较。结果表明，AExplainer 在三个数据集上都比基线方法取得了更好的表现，同时通过 t 检验分析，AExplainer 在 SST-2、Amazon 和 AGNews 数据集上与效果最好的基线模型的差异均具有显著性(p 值均小于 0.01)，表明 AExplainer 在小样本设置下能够显著提升分类性能。

值得注意的是，上面提到的基线模型是专门设计来提高问答任务准确性的，而 AExplainer 还专注于提供推理过程的解释。尽管关注点不同，但 AExplainer 不仅提供了对基础推理的洞察，而且在性能上也有所提升。

Table 1. The performance comparison of the ARExplainer model and various baseline methods on the SST-2, Amazon, and AGNews datasets under 1-shot setting
表 1. 在 1-shot 下 ARExplainer 模型与各种基线方法的在 SST-2、Amazon、AGNews 数据集性能比较

模型	SST-2	Amazon	AGNews
Fine-tuning	34.58	49.91	29.52
MV	80.43	88.72	75.31
AV	51.23	65.76	50.93
WARP	50.08	72.93	65.15
PV	54.91	69.68	67.39
ARExplainer	**87.17 ($p < 0.01$)**	**91.19 ($p < 0.01$)**	**79.40 ($p < 0.01$)**

4.5. 模型消融实验

表 2、表 3 总结了消融研究，针对 CommonsenseQA、Amazon 数据集检查了不同组件对模型的性能的影响。评估了不同 PLM 大小、数据增强机制和知识图谱推理引擎对数据集的影响。

4.5.1. PLM 大小

表 2 显示了 PLM 大小对本方法的影响。评估了三种不同 PLM 大小的性能：BERT-Base、BERT-Large 和 RoBERTa-large。结果表明，使用更大的 PLM 显著提高了性能。这些发现表明 PLM 的大小在模型的性能中发挥了关键作用，使用更大的 PLM 可以获得更好的性能。

4.5.2. 数据增强机制

本实验比较了 1-shot 的 Amazon 数据集，在不进行数据增强输入到知识图谱推理引擎和进行数据增强输入到推理引擎，由表 3 的结果可以得到，如果仅使用小样本对模型进行训练和推理，得到的效果远远差于进行数据增强后数据进行训练得到的结果。

Table 2. Performance comparison of the CommonsenseQA dataset across different models
表 2. CommonsenseQA 数据集在不同模型上的性能比较

方法	BERT-Base		BERT-Large		RoBERTTa-Large	
	Dev-Acc (%)	Test-Acc (%)	Dev-Acc (%)	Test-Acc (%)	Dev-Acc (%)	Test-Acc (%)
无知识图谱推理引擎	57.31	53.47	61.06	55.39	73.07	68.69
有知识图谱推理引擎	60.36	57.23	63.29	60.59	74.45	71.11

Table 3. Performance comparison of the Amazon dataset with and without data augmentation
表 3. Amazon 数据集在是否数据增强的性能比较

是否数据增强	Dev-Acc (%)	Test-Acc (%)
否	68.59	67.79
是	76.22	79.40

4.5.3. 知识图谱推理引擎

表 2 显示了知识图谱推理引擎对本方法的影响。比较了仅使用 LM 模型与使用来自 ConceptNet 的外

部知识的性能。仅模型意味着仅使用 LM 来预测答案。可以观察到，引入外部知识可以显著提高预测的准确性，对于不同的模型，引入外部知识可以都提高模型性能，至少可以提高 1.38%，这表明外部知识在增强模型的推理能力方面发挥了重要作用。

消融实验突显了本方法中每个组件的积极贡献。具体而言，PLM 的规模在整体性能提升中发挥了关键作用，知识图谱推理引擎的引入显著提升了模型的推理能力，而数据增强机制有效缓解了小样本学习的限制。

4.5.4. 计算复杂度分析

如表 4 所示，手动动词化器(MV)和自动动词化器(AV)中，|L|是标签数，k 是需要存储每个标签对应的候选词(或词 ID)；连续型动词化器：WARP，其中 T 是优化步数(通常较少)，|L|是标签数，d 是词嵌入维；连续型动词化器：原型动词化器(PV)，|L|是标签数，d 是词嵌入维；AREXplainer 在稀疏图(最大节点度 $\Delta \ll n$)上的时间复杂度和空间复杂度相对于最大路径长度 K 或节点数量 n 是线性的。

Table 4. Model complexity
表 4. 模型复杂度

模型	时间	空间
MV	$O(1)$	$O(\|L\|\cdot k)$
AV	$O(1)$	$O(\|L\|\cdot k)$
WARP	$O(T\cdot\|L\|\cdot d)$	$O(\|L\|\cdot d)$
PV	$O(\|L\|\cdot d)$	$O(\|L\|\cdot d)$
AREXplainer ($\Delta \ll n$)	$O(m^2n^2K\Delta)$	$O(mnK)$

Table 5. Explanation examples of AREXplainer, PathReasoner, and Explanations on the CommonsenseQA dataset
表 5. CommonsenseQA 数据集的 AREXplainer、PathReasoner 和 Explanations 的解释示例

模型输入	Q: What is someone doing if he or she is setting quietly and his or her eyes are moving? A. reading B. meditate C. fall asleep D. bunk E. think
标签	A. reading
PathReasoner	quietly [related to] quiet [at location] a library [used for] reading eyes [used for] reading eyes [form of] eye [related to] glasses [used for] reading sitting [related to] sit [related to] relaxing [has subevent] reading
Explanations for CommonsenseQA	**Positive examples:** When we read, our eyes move. While reading, a person sits quietly, **Negative examples:** While meditating, eyes don't move, eyes are closed, While sleeping, eyes are closed and they don’t move, When a person bunks, he/she doesn't sit quietly, Eyes don’t move when you think about **Explanation:** When we read, our eyes move. While reading, a person sits quietly. While meditating and sleeping, eyes don't move, eyes are closed. When a person bunks, he/she doesn't sit quietly. Eyes don’t move when you think about something.

续表

ARExplainer	当一个人安静地坐着，并且眼睛在移动时，根据图书馆通常用于阅读、眼睛的主要功能之一是阅读，可以得出他/她正在阅读(reading)是最合理的结论。

4.6. 实例的可解释性分析

表 2 的结果证明了 ARExplainer 的推理能力。为进一步验证其有效性，我们将其与两种先进方法 PathReasoner 和 CommonsenseQA 解释进行比较。PathReasoner 利用结构化信息生成推理路径，但其路径信息不完整，且需手动筛选，而 ARExplainer 不仅提供完整推理路径，还能生成合理的自然语言解释。如表 5 所示，PathReasoner 生成的路径存在冗余，难以识别真实推理过程，而 ARExplainer 则通过“为什么选择”解释，使推理过程更清晰可理解。

CommonsenseQA 解释数据集包含人工标注的解释，但其方法仅基于正负例的简单组合，未能体现模型的实际推理过程。相比之下，ARExplainer 通过逻辑推导生成解释，避免了仅靠句子组合的局限性，提供更完整、精准的推理过程描述，提升了可解释性和实用性。

5. 结论

在本文中，提出了 ARExplainer 方法，这是一种创新的模型，结合了数据增强、知识图谱推理和解释模块，旨在提升语言模型的性能，并提供清晰、可靠的推理解释。该模型有效解决了数据稀缺的问题，同时能够以逻辑且全面的方式解释推理结果，使得用户可以更容易通过自然语言理解模型的推理过程。实验结果表明，相较于先前的最先进方法，ARExplainer 在小样本数据集上表现出了优越的性能。分析表明，ARExplainer 不仅提升了模型的准确性，还显著增强了模型的可解释性，帮助用户更好地理解模型的决策过程。

6. 局限性

尽管 ARExplainer 在小样本文本分类中表现出优越性能与可解释性，但仍存在若干局限性。

知识图谱依赖：模型高度依赖外部知识图谱(如 ConceptNet)的质量与覆盖度。对于领域特定或语言小众的任务，现有通用图谱可能缺乏相关实体与关系，影响推理效果。

多阶段误差累积：ARExplainer 由数据增强、子图检索、多跳推理和解释生成四个阶段构成，各阶段误差可能逐级传递并放大，尤其在图谱链接不准确或 LLM 生成噪声较大时更为明显。

计算成本高昂：模型训练与推理涉及大规模图谱检索、多跳路径枚举和大语言模型 API 调用，导致时间与经济成本较高，在实时或低资源场景中的应用受限。

未来的研究可在轻量化推理、动态知识融合以及端到端的可解释结构等方面进一步探索，以克服上述局限。

基金项目

2025 年河北省硕士在读研究生创新能力培养资助项目(CXZZSS2025104)。

参考文献

[1] Bragg, J., Cohan, A., Lo, K., *et al*. (2021) Flex: Unifying Evaluation for Few-Shot NLP. *Advances in Neural Information Processing Systems* 34: *Annual Conference on Neural Information Processing Systems* 2021, *NeurIPS* 2021, 6-14 December 2021, 15787-15800.

[2] Schick, T., Schmid, H. and Schütze, H. (2020). Automatically Identifying Words That Can Serve as Labels for Few-Shot

Text Classification. *Proceedings of the* 28*th International Conference on Computational Linguistics*, Barcelona, December 2020, 5569-5578. https://doi.org/10.18653/v1/2020.coling-main.488

[3] Devlin, J., Chang, M.W., Lee, K., *et al*. (2019) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the* 2019 *Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, Volume 1, 4171-4186.

[4] Liu, Y. Ott, M., Goyal, N., *et al*. (2019) Roberta: A Robustly Optimized Bert Pretraining Approach.

[5] Raffel, C., Shazeer, N., Roberts, A., *et al*. (2020) Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, **21**, 1-67.

[6] Wei, J., Tay, Y., Bommasani, R., *et al*. (2022) Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research* (TMLR). https://doi.org/10.48550/arXiv.2206.07682

[7] Petroni, F., Rocktäschel, T., Riedel, S., Lewis, P., Bakhtin, A., Wu, Y., *et al*. (2019) Language Models as Knowledge Bases? *Proceedings of the* 2019 *Conference on Empirical Methods in Natural Language Processing and the* 9*th International Joint Conference on Natural Language Processing* (*EMNLP-IJCNLP*), Hong Kong, November 2019, 2463-2473. https://doi.org/10.18653/v1/d19-1250

[8] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., *et al*. (2023) Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, **55**, 1-38. https://doi.org/10.1145/3571730

[9] Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., *et al*. (2023) A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. *Proceedings of the* 13*th International Joint Conference on Natural Language Processing and the* 3*rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, Volume 1, 675-718. https://doi.org/10.18653/v1/2023.ijcnlp-main.45

[10] Zhang, H., Song, H., Li, S., Zhou, M. and Song, D. (2023) A Survey of Controllable Text Generation Using Transformer-Based Pre-Trained Language Models. *ACM Computing Surveys*, **56**, 1-37. https://doi.org/10.1145/3617680

[11] Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B. and Sen, P. (2020) A Survey of the State of Explainable AI for Natural Language Processing. *Proceedings of the* 1*st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the* 10*th International Joint Conference on Natural Language Processing*, Suzhou, December 2020, 447-459. https://doi.org/10.18653/v1/2020.aacl-main.46

[12] Situ, X., Zukerman, I., Paris, C., Maruf, S. and Haffari, G. (2021) Learning to Explain: Generating Stable Explanations Fast. *Proceedings of the* 59*th Annual Meeting of the Association for Computational Linguistics and the* 11*th International Joint Conference on Natural Language Processing*, Volume 1, 5340-5355. https://doi.org/10.18653/v1/2021.acl-long.415

[13] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H. and Neubig, G. (2023) Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, **55**, 1-35. https://doi.org/10.1145/3560815

[14] Chen, Q., Ji, F., Zeng, X., Li, F., Zhang, J., Chen, H., *et al*. (2021) KACE: Generating Knowledge Aware Contrastive Explanations for Natural Language Inference. *Proceedings of the* 59*th Annual Meeting of the Association for Computational Linguistics and the* 11*th International Joint Conference on Natural Language Processing*, Volume 1, 2516-2527. https://doi.org/10.18653/v1/2021.acl-long.196

[15] Abu-Salih, B. (2021) Domain-Specific Knowledge Graphs: A Survey. *Journal of Network and Computer Applications*, **185**, Article ID: 103076. https://doi.org/10.1016/j.jnca.2021.103076

[16] Chen, J., Tam, D., Raffel, C., Bansal, M. and Yang, D. (2023) An Empirical Survey of Data Augmentation for Limited Data Learning in NLP. *Transactions of the Association for Computational Linguistics*, **11**, 191-211. https://doi.org/10.1162/tacl_a_00542

[17] Møller, A.G., Dalsgaard, J.A., Pera, A., *et al*. (2023) Is a Prompt and a Few Samples All You Need? Using GPT-4 for Data Augmentation in Low-Resource Classification Tasks.

[18] Shum, K., Diao, S. and Zhang, T. (2023) Automatic Prompt Augmentation and Selection with Chain-of-Thought from Labeled Data. *Findings of the Association for Computational Linguistics*: *EMNLP* 2023, Singapore, December 2023, 12113-12139. https://doi.org/10.18653/v1/2023.findings-emnlp.811

[19] Peng, B., Li, C., He, P., *et al*. (2023) Instruction Tuning with GPT-4.

[20] Hoffmann, J., Borgeaud, S., Mensch, A., *et al*. (2022) Training Compute-Optimal Large Language Models. *Proceedings of the* 36*th International Conference on Neural Information Processing Systems*, New Orleans, 28 November-9 December 2022, 30016-30030.

[21] Luo, L., Zhao, Z., Haffari, G., *et al*. (2024) Graph-Constrained Reasoning: Faithful Reasoning on Knowledge Graphs with Large Language Models. 42*nd International Conference on Machine Learning*. https://doi.org/10.48550/arXiv.2410.13080

[22] Sun, Y., Wang, S., Feng, S., *et al*. (2021) Ernie 3.0: Large-Scale Knowledge Enhanced Pre-Training for Language Understanding and Generation.

[23] Schick, T. and Schütze, H. (2021) Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 255-269. https://doi.org/10.18653/v1/2021.eacl-main.20

[24] Yuan, W., Neubig, G. and Liu, P. (2021) Bartscore: Evaluating Generated Text as Text Generation. *Advances in Neural Information Processing Systems*, **34**, 27263-27277.

[25] Haviv, A., Berant, J. and Globerson, A. (2021) BERTese: Learning to Speak to BERT. *Proceedings of the* 16*th Conference of the European Chapter of the Association for Computational Linguistics*: *Main Volume*, April 2021, 3618-3623. https://doi.org/10.18653/v1/2021.eacl-main.316

[26] Li, X.L. and Liang, P. (2021) Prefix-Tuning: Optimizing Continuous Prompts for Generation. *Proceedings of the* 59*th Annual Meeting of the Association for Computational Linguistics and the* 11*th International Joint Conference on Natural Language Processing*, Volume 1, 4582-4597. https://doi.org/10.18653/v1/2021.acl-long.353

[27] Tsimpoukelli, M., Menick, J.L., Cabi, S., *et al*. (2021) Multimodal Few-Shot Learning with Frozen Language Models. *Advances in Neural Information Processing Systems* 34: *Annual Conference on Neural Information Processing Systems* 2021, *NeurIPS* 2021, 6-14 December 2021, 200-212.

[28] Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M. and Liu, Q. (2019). ERNIE: Enhanced Language Representation with Informative Entities. *Proceedings of the* 57*th Annual Meeting of the Association for Computational Linguistics*, Florence, July 2019, 1441-1451. https://doi.org/10.18653/v1/p19-1139

[29] Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., *et al*. (2020) K-BERT: Enabling Language Representation with Knowledge Graph. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 2901-2908. https://doi.org/10.1609/aaai.v34i03.5681

[30] Liu, Y., Wan, Y., He, L., Peng, H. and Yu, P.S. (2021) KG-BART: Knowledge Graph-Augmented BART for Generative Commonsense Reasoning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 6418-6425. https://doi.org/10.1609/aaai.v35i7.16796

[31] Lin, B.Y., Chen, X., Chen, J. and Ren, X. (2019) Kagnet: Knowledge-Aware Graph Networks for Commonsense Reasoning. *Proceedings of the* 2019 *Conference on Empirical Methods in Natural Language Processing and the* 9*th International Joint Conference on Natural Language Processing* (*EMNLP-IJCNLP*), Hong Kong, November 2019, 2829-2839. https://doi.org/10.18653/v1/d19-1282

[32] Dai, D., Dong, L., Hao, Y., Sui, Z., Chang, B. and Wei, F. (2022) Knowledge Neurons in Pretrained Transformers. *Proceedings of the* 60*th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 8493-8502. https://doi.org/10.18653/v1/2022.acl-long.581

[33] Rosset, C., Xiong, C., Phan, M., *et al*. (2020) Knowledge-Aware Language Model Pretraining.

[34] Lewis, P., Perez, E., Piktus, A., *et al*. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems* 33: *Annual Conference on Neural Information Processing Systems* 2020, *NeurIPS* 2020, 6-12 December 2020, 9459-9474.

[35] Wang, X., Kapanipathi, P., Musa, R., Yu, M., Talamadupula, K., Abdelaziz, I., *et al*. (2019) Improving Natural Language Inference Using External Knowledge in the Science Questions Domain. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 7208-7215. https://doi.org/10.1609/aaai.v33i01.33017208

[36] Hambardzumyan, K., Khachatrian, H. and May, J. (2021) WARP: Word-Level Adversarial Reprogramming. *Proceedings of the* 59*th Annual Meeting of the Association for Computational Linguistics and the* 11*th International Joint Conference on Natural Language Processing*, Volume 1, 4921-4933. https://doi.org/10.18653/v1/2021.acl-long.381

[37] Cui, G., Hu, S., Ding, N., Huang, L. and Liu, Z. (2022) Prototypical Verbalizer for Prompt-Based Few-Shot Tuning. *Proceedings of the* 60*th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 7014-7024. https://doi.org/10.18653/v1/2022.acl-long.483