

# 효율적인 다국어 지식 공유를 통한 다국어 지식 그래프 완성

Cunli Mao<sup>1,2</sup>, Xiaofei Gao<sup>1,2</sup>, Ran Song<sup>1,2\*</sup>, Shizhu He<sup>3,4</sup>  
 Shengxiang Gao<sup>1,2</sup>, Kang Liu<sup>3,4</sup>, Zhengtao Yu<sup>1,2<sup>1</sup></sup> 정보 공학 및 자동  
 화 학부,

중국 쿠밍, 쿠밍 과학기술대학교

<sup>2</sup> 중국 원난성 인공지능 핵심 연구소, 중국 쿠밍

<sup>3</sup> 중국과학원 자동화연구소, 복잡계 인지 및 의사결정 지능 핵심 연구소, 중국 베이징

<sup>4</sup> 중국과학원대학 인공지능학부, 중국 베이징

{maocunli,xiaofeigao\_g,song\_ransr}@163.com, {shizhu.he,kliu}@nlpr.ia.ac.cn,  
 {g aoshengxiang.yn,ztyu}@hotmail.com

## 초록

대규모 언어 모델(LLM) 기반 다국어 지식 그래프 완성(MKGC)은 LLM의 다국어 이해 능력을 활용하여 누락된 사실을 예측함으로써 다국어 지식 그래프(KG)의 완성도를 향상시키는 것을 목표로 합니다. 그러나 기존 MKGC 연구는 LLM의 다국어 능력을 충분히 활용하지 못하며, 언어 간 지식의 공유 가능성을 간과하고 있습니다. 본 논문에서는 다국어 공유 지식을 활용하는 새로운 MKGC 프레임워크를 제안합니다.

두 가지 구성 요소인 지식 수준 그룹화 전문가 혼합 모델(KL-GMoE)과 반복적 엔티티 재정렬(IER)을 통해 성능을 크게 향상시킵니다. KL-GMoE는 공유 지식을 효율적으로 모델링하는 반면, IER는 그 활용도를 크게 높입니다. 본 프레임워크를 평가하기 위해 5개 언어로 구성된 mKG 데이터셋을 구축하고 비교 분석을 수행했습니다.

기존 최첨단(SOTA) MKGC 방법과의 포괄적인 비교 실험을 수행하였다. 실험 결과는 본 프레임워크가 SOTA MKGC 방법 대비 Hits@1, Hits@3, Hits@10 지표에서 각각 5.47%, 3.27%, 1.01%의 성능 향상을 달성함을 보여준다. 추가 실험 분석을 통해 미관측 언어 및 불균형 언어 환경에서의 지식 공유 특성을 규명하였습니다. 본 연구의 데이터셋과 코드는 <https://github.com/gaoxiaofei07/KL-GMoE>에서 공개하였습니다.

## 1 소개

지식 그래프(KG)(Weikum, 2021)는 현실 세계에 대한 지식을 표현하고 조직화하기 위해 설계된 구조화된 의미적 지식 기반입니다. 대부분의 KG는 위키데이터(Wikidata)(Vrandecic' and Krötzsch,

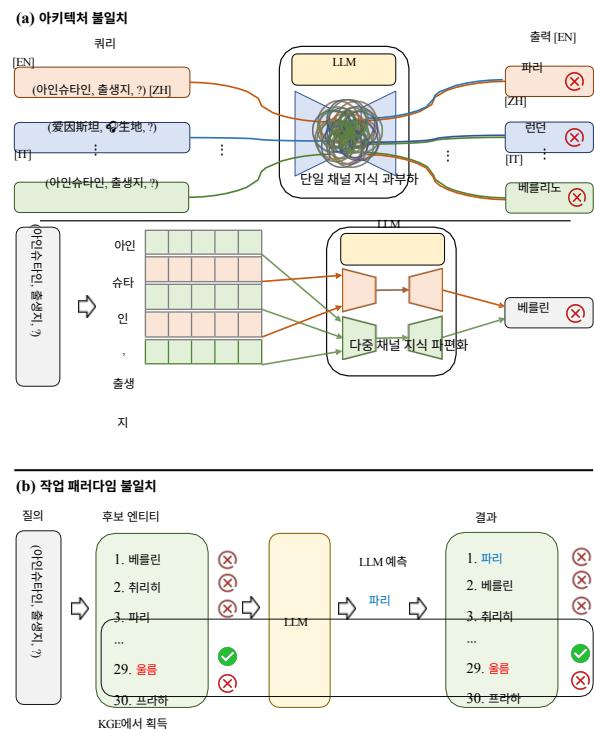


그림 1: 이 그림은 발생하는 문제들을 보여줍니다

LLM을 MKGC 작업에 직접 적용할 때. (a)는 기존 PEFT가 MKGC 작업에 적합하지 않음을 보여준다. (b)는 LLM이 탁월한 작업 패러다임과 MKGC 작업 간의 불일치를 나타낸다.

기존 다국어 지식 그래프(KG)는 종종 불완전하여 실용적 적용에서 효과성이 제한된다(Ji et al., 2022). 다국어 지식 그래프 완성(MKGC)은 알려진 다국어 지식을 활용하여 누락된 삼중화를 완성하고 KG의 완전성을 향상시키는 것을 목표로 한다.

기존 연구는 엔티티와 관계를 저차원 벡터 공간으로 매핑하여 완성을 달성하는 임베딩 기반 방법(Ge et al., 2024)에 집중해왔다. 최근 언어 모델의 발전으로 MKGC 연구는 KG 완성을 재구성하는 생성 기반 접근법(Chen et al., 2022; Saxena et al., 2022)으로 전환되고 있다.

\*교신저자

텍스트 생성을 위한 방법론으로 제시된다. (Song et al., 2023; Zhou et al., 2022)는 단일 사전 훈련된 언어 모델(PLM)을 활용하여 통합된 의미 공간 내에서 다국어 지식을 통합함으로써 MKGC에서 우수한 성능을 달성한다. 또한 DIFT(Liu et al., 2024)와 같은 최근 연구는 LLM의 미세 조정을 통한 작업 적응을 탐구하여 단일 언어 KGC 작업에서 강력한 성능을 보여준다. 다양한 코퍼스로 사전 훈련된 현대 LLM은 본질적으로 다국어 능력을 보유하며(Huang et al., 2024), 통합 모델 내에서 언어 간 표현 및 지식 공유를 가능하게 합니다(Hu et al., 2025). 특히, 이러한 내재적 다국어 지식 공유 능력은 MKGC에 필수적이며 완성 성능 향상에 상당한 잠재력을 제공합니다. 이러한 잠재력을 인식하여 본 연구는 LLM의 내재적 능력을 활용해 MKGC를 개선하는 효과적인 방법을 탐구한다.

그러나 LLM을 MKGC에 직접 적용하는 것은 주로 두 가지 핵심 측면, 즉 모델 아키텍처와 작업 패러다임에서 비롯된 여러 가지 과제를 안고 있습니다. 1) **아키텍처 불일치**: LLM을 위한 기존의 파라미터 효율적 미세 조정(PEFT) 방법(Han et al., 2024)은 주로 텍스트 중심 작업을 위해 설계되었으며 지식 수준 작업에 적용할 때 상당한 격차를 보입니다. 특히 단일 채널 방식은 KG의 복잡한 다국어 특성에 취약합니다. 그림 1(a) 상단에서 보듯, 다수의 다국어 질의를 단일 채널로 처리하면 지식 과부하가 발생합니다. 이 과부하는 모델이 언어 간 유사한 지식을 이해하는 능력을 저해하여 잘못된 예측을 초래합니다. 예를 들어, *아인슈타인의 출생지에* 관한 영어, 중국어, 이탈리아어 질의 모두 잘못된 결과를 산출합니다. 반대로 다중 채널 방식은 지식의 원자성을 훼손하여 지식 분열을 유발하는 경향이 있습니다. 그림 1(a) 하단에서 볼 수 있듯이, 쿼리 토큰이 서로 다른 채널에서 처리됩니다. 이러한 분열된 처리는 결국 잘못된 엔티티 예측으로 이어집니다.

2) **작업 패러다임 불일치**: MKGC 작업은 엔티티 순위를 매기는 작업으로, 텍스트 생성 패러다임과 불일치를 보입니다. 그림 1(b)에서 볼 수 있듯이, 쿼리 (*아인슈타인 출생지*, ?), LLM은 *파리를* 오답으로 예측했습니다. 이 선택은 정답 *엔티티/인 물체의 순위*를 개선하지 못했습니다.

1) **아키텍처 불일치**, 특히 지식 과부하 문제를 해결하기 위해 전용 지식 채널 수를 증가시키는 방안을 제안한다.

네스. 이를 통해 각 채널은 의미적으로 유사한 정보 처리에 집중할 수 있어, LLM의 다국어 공유 지식 이해 및 활용 능력을 향상시킵니다. 동시에 각 채널이 완전한 지식을 독립적으로 처리할 수 있도록 함으로써 지식 분열을 효과적으로 완화하고 모델의 다국어 정보에 대한 포괄적 이해를 촉진합니다. 2) **작업 패러다임 불일치**를 해결하기 위해, 우리는 LLM의 훈련 목표를 조정하여 다중 엔티티의 순위를 반복적으로 정교화할 수 있도록 제안합니다. 이 접근법은 LLM이 다국어 공유 지식을 활용하는 빈도를 높여 올바른 엔티티의 순위를 향상시키는 것을 목표로 합니다.

본 논문에서는 다국어 공유 지식을 효과적으로 활용하여 MKGC의 성능을 향상시키는 새로운 프레임워크를 제안한다. 제안된 프레임워크는 지식 수준 그룹화 전문가 혼합(KL-GMoE)과 반복적 엔티티 재정렬(IER)이라는 두 가지 시너지 구성 요소로 이루어져 있다. KL-GMoE는 지식 수준 전문가 라우팅 메커니즘과 그룹 기반 전문가 혼합(MoE) 아키텍처를 도입한다. 이 설계는 지식 분열을 완화하는 동시에 LLM의 다국어 공유 지식 포착 능력을 크게 향상시키는 것을 목표로 한다. IER는 LLM의 훈련 목표와 디코딩 전략을 모두 수정합니다. 이를 통해 모델은 다중 반복 정제를 통해 다국어 공유 지식 활용 능력을 크게 향상시킬 수 있습니다. 실험 결과, 본 프레임워크는 SOTA MKGC 방법에 비해 Hits@1, Hits@3, Hits@10 지표에서 각각 5.47%, 3.27%, 1.01%의 개선을 달성한 것으로 나타났습니다. 추가 실험 분석을 통해 미지 언어 및 불균형 언어 환경에서의 지식 공유 특성을 규명하였습니다.

요약하면, 우리의 기여는 다음과 같습니다:

- 모델 아키텍처 불일치 문제를 해결하고 공유 지식을 효율적으로 모델링하기 위해 KL-GMoE를 제안합니다.
- 작업 패러다임의 불일치를 해결하기 위해 IER를 제안하여 공유 지식 활용도를 높였습니다.
- 실험 결과, 우리의 프레임워크는 SOTA MKGC 방법보다 평균적으로 Hits@1, Hits@3, Hits@10에서 각각 5.47%, 3.27%, 1.01% 향상된 성능을 보이며 현저히 우수한 성능을 보였습니다.

언어 엔티티 관계 훈련 검증 테스트					
EN	86,539	512	708,267	49,782	49,777
FR	89,754	478	839,623	49,908	30,000
IT	65,434	445	613,014	49,883	20,000
JA	46,294	432	321,237	49,939	10,000
ZH	63,278	397	546,626	49,969	10,000
합계	351,299	2,264	3,028,767,249	481,119,777	

표 1: 다국어 지식 그래프 완성 데이터 세트의 통계.

## 2 데이터 세트

### 2.1 데이터셋 구축

우리는 위키데이터와 위키백과를 통합한 백만 규모 영어 지식 그래프(KG) 데이터셋인 위키데어/E5M(Wang et al., 2021)을 기초 시드 라이브러리로 활용합니다. 이를 바탕으로 위키데이터에서 데이터를 수집하여 프랑스어, 이탈리아어, 중국어, 일본어를 추가로 포함하도록 데이터셋을 확장했습니다. 표 1에 제시된 바와 같이, 엔티티 수, 관계 수, 훈련용, 검증용, 테스트용 트리플 수에 대한 통계를 제시합니다. 이 지식 그래프는 351,299개의 엔티티와 2,264개의 관계를 포함하며, 총 트리플 수는 300만 개를 초과합니다.

다언어 지식 분포의 특성에 기반하여, 서로 다른 언어 간 지식은 완전히 정렬되어 있지 않으며 특정 언어적 특수성을 나타낸다(Song et al., 2025). 이러한 언어 간 지식의 비대칭성은 일부 지식이 특정 언어에 국한됨을 시사한다. 따라서 우리가 구축한 데이터셋은 지식의 자연스러운 분포 패턴을 따른다. 일부 지식은 여러 언어에 걸쳐 공유되며, 이는 언어 간의 유사성을 반영한다. 다른 지식은 각 언어에 고유하며, 이는 각 언어의 독특한 특성을 반영한다.

### 2.2 프롬프트 구성

우리는 DIFT(Liu et al., 2024)에서 제안한 프롬프트 구성 방법을 채택합니다. 임베딩 기반 모델은 훈련 데이터를 학습했기 때문에 대부분의 훈련 사실에 대해 후보 엔티티 중 정답 엔티티를 첫 번째 순위로 배치하는 경향이 있습니다. 이러한 순위 지정된 후보를 사용하여 프롬프트를 구성하면 LLM이 첫 번째 엔티티를 정답으로 선택하는 편향을 발달시킬 수 있습니다. 따라서 검증 세트에서 일부를 분할하여 프롬프트를 구성하고, 이를 LLM의 미세 조정 단계에서 훈련 데이터로 활용합니다.

쿼리  $q = (h, r, ?)$ 에 대해, 구성된 프롬프트  $P$ 는 네 부분으로 구성됩니다: 쿼리  $Q$ , 설명자  $D$ , 인접 사실  $N$ , 후보 엔티티  $M_c$ . 이는 다음과 같이 표현될 수 있습니다:

, 인접 사실  $N$ , 후보 엔티티  $M_c$ 로 구성됩니다. 이는 다음과 같이 표현될 수 있습니다:

$$P(q) = [Q; D; N; M_c]. \quad (1)$$

설명은  $h$ 에 대한 구체적인 설명 정보를 제공하여 모델이 엔티티의 의미를 보다 정확하게 이해할 수 있도록 합니다. 인접 사실은 엔티티  $h$ 를 포함하는 삼중항으로, 이러한 삼중항은 지식 그래프 임베딩(KGE) 모델의 훈련 데이터에서 무작위로 추출됩니다. 이러한 인접 사실들은 LLM이  $h$ 에 대한 이해를 향상시키기 위한 것입니다. 후보 엔티티  $M_c = [e_1, e_2, \dots, e_m]$ 은 KGE 모델(Bordes et al., 2013)이 생성한 순위 결과에서 상위  $m$  개 엔티티를 선택하여 구성됩니다. LLM이 MKGC의 작업 패리다임에 적응할 수 있도록, 훈련 과정에서  $M_c$  내 엔티티 수  $m$  을 처리했습니다. 구체적인 처리 방법은 3.3절에 상세히 기술되어 있습니다. 부록 A.1에는 구체적인 프롬프트 예시를 제공합니다.

## 3 방법론

### 3.1 작업 정의

본 논문에서는 KGE 모델과 LLM을 통합하여 MKGC 작업을 수행한다. 먼저, 쿼리  $q = (h, r, ?)$ 에 대해 KGE 모델을 사용하여 상위  $m$  개 순위의 엔티티를 추출하며, 이는 후보 엔티티 집합  $M_{(c)}$  형성한다. 다음으로, LLM을 활용하여 후보 엔티티 집합  $M_c$ 에서 최적의 엔티티를 선택하여 쿼리  $q$ 를 완성한다. 완성 과정은 다음과 같이 공식화할 수 있다:

$$\hat{e} = \operatorname{argmax}_{e_i \in M_c} P(e_i | h, r, M_c), \quad (2)$$

여기서  $\hat{e}$ 는 쿼리  $q$ 를 완성하는 최적의 엔티티를 나타내며,  $P(e_i | h, r, M_c)$ 는 헤드 엔티티  $h$ , 관계  $r$  및 후보 엔티티 집합  $M_c$ 가 주어졌을 때 엔티티  $e_i$ 가 선택될 확률을 나타낸다.

엔티티  $h$ , 관계  $r$  및 후보 엔티티  $M_c$ 를 주어진 상태에서 엔티티  $e_i$ 를 선택할 확률을 나타냅니다.

### 3.2 KL-GMoE 아키텍처

KL-GMoE는 MKGC 작업에 특화되어 설계되었습니다. 이 아키텍처는 단일 채널로 인한 지식 과부하를 완화하고 LLM의 공유 지식 포착 능력을 향상시키기 위해 다중 전문가 그룹으로 구성됩니다. 또한 KL-GMoE는 지식 수준 전문가가 라우팅 메커니즘을 활용하여 모든 전문가가 공동으로 참여하는 대신 각 샘플이 특정 전문가에 의해 처리되도록 보장합니다. 그림 2(a)에서 볼 수 있듯이,

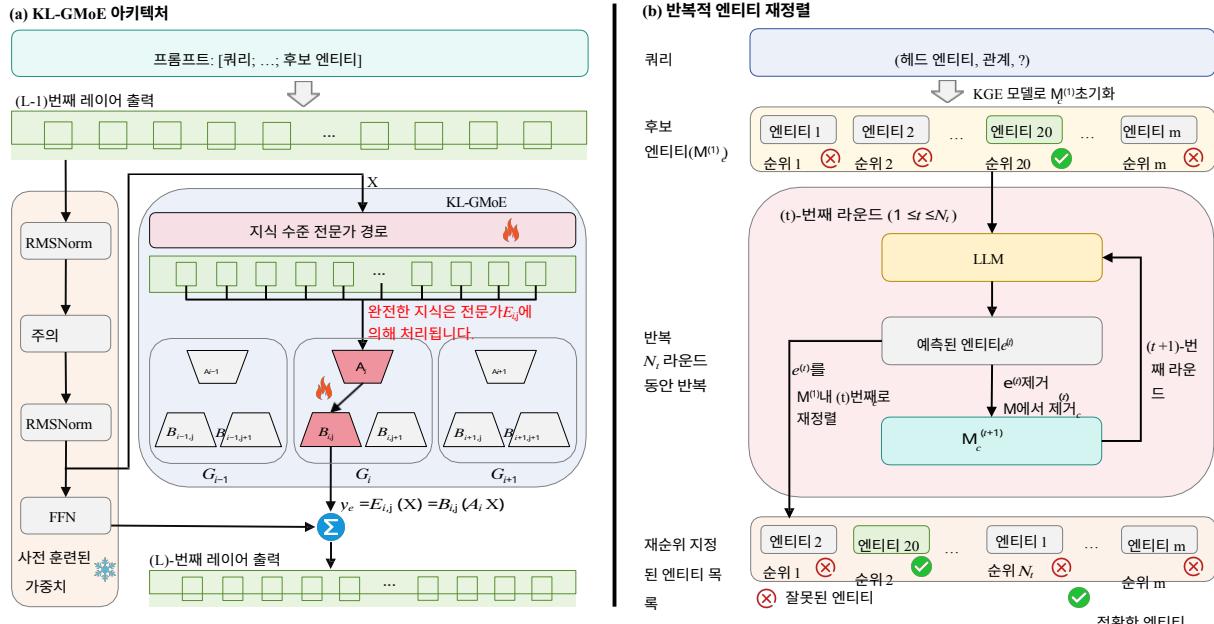


그림 2: 본 연구에서 제안한 프레임워크를 보여줍니다. 그림 (a)는 KL-GMoE의 아키텍처와 워크플로우를 나타내며, **빨간색으로 강조된 행렬**  $A_i$ 와  $B_{ij}$ 는 현재 활성화된 전문가를 나타냅니다. 그림 (b)는 IER 방법의 워크플로우를 보여줍니다.  $N_r$  번의 반복 후, 우리는 재정렬된 엔티티 목록을 얻을 수 있습니다.

KL-GMoE로 처리되는 각 샘플에 대해, 전문가 그룹  $G_i$ 내의 행렬  $A_i$ 와 하나의 행렬  $B_{ij}$ 만이 활성화됩니다. KL-GMoE는 LLM의 전방전파 네트워크(FFN) 계층에만 적용됩니다. 구체적으로, 전방전파 중 FFN 계층에서의 행렬 연산은 다음과 같이 표현될 수 있습니다:

$$\mathbf{y} = \mathbf{W}_0 \mathbf{X} + \mathbf{y}_e, \quad (3)$$

$\mathbf{W}_0 \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$ 은 훈련 중 고정되는 원래 FFN 레이어 매개 변수 행렬을 나타낸다.

$\mathbf{X} = [\mathbf{x}_h; \mathbf{x}_r; \mathbf{x}_l]$ 는 FFN 레이어의 입력을 나타냅니다.  $\mathbf{y}_e$ 는 입력  $\mathbf{X}$ 를 기반으로 KL-GMoE가 계산한 출력을 나타냅니다.

KL-GMoE의 설계는 Hy-draLoRA(Tian et al., 2024)에서 제안된 비대칭 지표 미세 조정 아키텍처에서 영감을 받았습니다. 미터법 미세 조정 아키텍처에서 영감을 받았습니다. 우리는 그룹화

MoE 설계 아키텍처에서 각 그룹은 다음과 같이 표현될 수 있습니다:

$$G_i = (A_i, \{B_{ij} \mid j \in \{1, 2, \dots, N_b\}\}), \quad (4)$$

여기서  $i \in \{1, 2, \dots, N_g\}$ ,  $N_g$  는 총 전문가 그룹 수를 나타낸다.  $N_b$  는 B의 수이다.

군  $G_i$ 내의 행렬들. 각 군  $G_i$ 내에서,  $A_i$ 행렬과  $B_{ij}$ 행렬의 짙짓기는 전문가  $E_{ij} = (A_i, B_{ij})$ 로 간주된다.  $A_i$ 행렬은 유사한 지식 범주를 포착하도록 설계되었다. 그룹  $G_i$ 내의 서로 다른  $B_{ij}$ 행렬들은 이 지식 범주 내 미묘한 차이를 포착하는 모듈

다중 언어 간 공유된 지식을 포착하는 LLM의 능력을 향상 시키는 것을 목표로 합니다.

동시에, 제안된 지식 수준 전문가 라우팅 메커니즘은 세 가지 서로 다른 경로인  $\mathbf{R}_g$ ,  $\mathbf{R}_k$  및  $\mathbf{R}_{(l)}$ 를 포함합니다. 먼저  $\mathbf{R}_g$ 에 기반하여 전문가 그룹이 선택됩니다. 이 그룹 내에서  $\mathbf{R}_k$ 와  $\mathbf{R}_{(l)}$ 를 결합하여 특정 전문가가 결정됩니다. 다음은 이 세 가지 경로를 기반으로 특정 전문가를 선택하는 과정을 설명합니다.

$\mathbf{R}_g$ 는  $\mathbf{X}$ 를 처리할 전문가 그룹을 결정하는 그룹 라우팅 선택 모듈입니다. 그룹 선택은 다음과 같이 공식화됩니다:

$$G_i = \underset{i \in \{1, 2, \dots, N_g\}}{\operatorname{argmax}} (\mathbf{R}_g(\mathbf{X}))$$

$$= \underset{i \in \{1, 2, \dots, N_g\}}{\operatorname{argmax}} \sum_{m \in \{h, r, t\}} \operatorname{Softmax}(\mathbf{W}_g \mathbf{x}_m), \quad (5)$$

로 간주된다. 이러한 설계는

여기서  $\mathbf{W}_g \in \mathbb{R}^{Ng \times din}$ 은  
 $G_i$ 는 X 처리를 위해 선택된 전문가 그룹을 나타냅니다.  
X를 처리하도록 선택된 그룹을 나타냅니다.

$\mathbf{R}_k$  및  $\mathbf{R}_i$ 는 그룹  $G_i$ 내에서 작동하는 전문가 라우팅 선택 모듈을 나타냅니다. 이 모듈들은 입력 X와 A, 행렬의 출력을 종합적으로 고려하여 전문가 선택을 수행합니다. 구체적으로,  $\mathbf{R}_{(k)}$ 는 X를 기반으로 전문가 선택 점수  $S_k$ 를 생성합니다. 공식은 다음과 같습니다.

$\mathbf{S}_k$ 의 계산은 다음과 같다:

$$\mathbf{S}_k = \mathbf{R}_k(\mathbf{X}) = \sum_{m \in \{h, r, t\}} \text{Softmax}(\mathbf{W}_k \mathbf{x}_m), \quad (6)$$

여기서  $\mathbf{S}_k \in \mathbb{R}^{Nb}$ , 그리고  $\mathbf{W}_k \in \mathbb{R}^{Nb \times \text{dim}}$ 은  $\mathbf{X}$ 를 입력으로 받는 라우팅 행렬이다.  $\mathbf{R}_k$ 은 행렬  $A(i)$ 의 출력  $A(i)\mathbf{X}$ 를 기반으로 전문가 선택 점수  $S(I)$ 를 생성한다.  $S(I)$ 의 계산은 다음과 같이 표현될 수 있다:

전문가 선택 점수  $S(I)$ 를 생성하며,  $I$ 는 행렬  $A_i$ 의 출력  $A_i \mathbf{X}$ 에 기반합니다.  $S_I$ 의 계산은 다음과 같이 표현될 수 있습니다:

$$\mathbf{S}_I = \mathbf{R}_I(\mathbf{X}) = \sum_{m \in \{h, r, t\}} \text{Softmax}(\mathbf{W}_I(A_i \mathbf{x}_m)), \quad (7)$$

여기서  $\mathbf{S}_I \in \mathbb{R}^{Nb}$ 이며,  $\mathbf{W}_I \in \mathbb{R}^{Nb \times r}$ 은 입력으로  $A_i \mathbf{X}$ 를 받는 라우팅 행렬이다.  $r$ 은

LoRA(Hu et al., 2022)에서 순위의 크기를 나타냅니다. 그런 다음,  $\mathbf{X}$ 를 처리하기 위해  $\mathbf{S}_k$ 와  $\mathbf{S}_I$ 의 점수에 기반하여 그 률  $G_i$ 에서 행렬  $B_{i,j}$ 를 선택합니다:

$$B_{i,j} = \underset{j \in \{1, 2, \dots, N_b\}}{\operatorname{argmax}} (\mathbf{S}_k + \mathbf{S}_I). \quad (8)$$

마지막으로, 지식 수준 전문가 라우팅 메커니즘에 기반하여 전문가  $E_{i,j} = (A_i, B_{i,j})$ 가  $\mathbf{X}$ 를 처리한다고 결정합니다.

전문가  $E_{i,j}$ 를 결정한 후, KL-GMoE의 출력은 다음과 같이 표현됩니다:

$$\mathbf{y}_e = E_{i,j}(\mathbf{X}) = B_{i,j}(A_i \mathbf{X}). \quad (9)$$

그런 다음, 방정식 3에 표시된 바와 같이 전문가 출력  $\mathbf{y}_e$ 가 원래 FFN 출력에 추가됩니다.

### 3.3 반복적 엔티티 재정렬

우리는 LLM의 교차 언어 공유 지식 활용을 향상시키기 위한 반복적 엔티티 재정렬(IER)이라는 방법을 제안한다. 그림 2(b)에서 볼 수 있듯이, IER 방법은 여러 번의 반복을 통해 공유 지식을 완전히 활용하여 올바른 엔티티 순위의 정확도를 크게 향상시킨다. IER는 LLM의 훈련 과제와 디코딩 전략을 조정합니다. 훈련 단계에서 후보 엔티티 수  $m$ 을 변수 값으로 무작위 설정하여 LLM이 다중 엔티티 순위를 반복적으로 조정할 수 있도록 훈련합니다. 디코딩 단계에서는 IER가 LLM이 다중 엔티티 예측을 여러 차례 수행하여 순위를 조정할 수 있도록 합니다.

쿼리  $q = (h, r, ?)$ 에 대해, 초기 후보 엔티티 집합은 KGE 모델에 의해 생성되며  $M^{(1)} = [e_1, e_2, \dots, e_m]$ 로 표기된다. 엔티티 목록은

정렬 대상 객체는  $L^{(1)} = M^{(1)}$ 로 초기화된다.

t-번째 라운드에서, 여기서  $t \in \{1, 2, \dots, N_r\}$ , 엔티티 예측 연산은 다음과 같이 표현될 수 있다:

$$e^{(t)} = \underset{e_t \in M^{(t)}}{\operatorname{argmax}} P(e_t | h, r, M^{(t)}), \quad (10)$$

$M^{(t)}$ 는 라운드  $t$ 에서의 후보 엔티티 집합을 나타낸다.  $e^{(t)} \in M^{(t)}$ 로부터 LLM이 예측한 엔티티이다. 그런 다음 다음 반복을 위해  $M^{(t)}$ 를 업데이트하여  $M^{(t+1)}$ 를 얻습니다:

$$M^{(t+1)} = M^{(t)} \setminus \{e^{(t)}\}, \quad (11)$$

$M^{(t)} \setminus \{e^{(t)}\}$ 는  $M^{(t)}$ 에서 엔티티  $e^{(t)}$ 를 제거함을 나타낸다. 마지막으로,  $L^{(t)}$  내 엔티티  $e^{(t)}$ 의 순위를 업데이트 합니다:

$$L^{(t+1)} = \text{Insert}(L^{(t)} \setminus \{e^{(t)}\}, t, e^{(t)}), \quad (12)$$

where  $\text{Insert}(L^{(t)} \setminus \{e^{(t)}\}, t, e^{(t)})$ 는 먼저  $L^{(t)}$ 에서  $e^{(t)}$ 를 제거한 후,  $e^{(t)}$ 를  $L^{(t)}$ 의 t번째 위치에 삽입하는 것을 의미합니다.  $N_r$ 번 반복 한 후 라운드를 거친 후, 최종 순위 목록  $L^{(N_r+1)}$ 을 얻습니다. IER 구현은 부록 A.2에 상세히 설명되어 있습니다.

## 4 실험

### 4.1 구현 세부 사항

실험에서 후보 엔티티를 추출하기 위해 TransE(Bordes et al., 2013)를 선택했습니다. 또한 파인 튜닝을 위한 베이스 모델로 Llama-2-7b-chat-hf<sup>1</sup>을 활용했습니다. 모델 훈련 하이퍼파라미터는 다음과 같이 설정되었습니다: 학습률은 2e-5, LoRA 랭크는 4, 후보 엔티티 길이  $M_{(c)}$ 는 25~30 사이입니다. IER의 반복 횟수  $N_r$ 는 10입니다.

### 4.2 다국어 지식 그래프 완성

구축한 데이터셋에서 제안된 프레임워크의 성능을 임베딩 기반 및 생성 기반 방법과 비교했습니다. 실험 결과는 제안된 방법이 모든 언어에 걸쳐 평균 메트릭에서 최적의 성능을 달성함을 보여줍니다. 구체적으로, 표 2에서 볼 수 있듯이, 제안된 프레임워크의 성능은 EN, FR, IT 세 언어에서 앞서 언급된 모든 방법을 능가합니다. JA와 ZH의 경우, Hits@10을 제외한 모든 지표에서 우수한 성능을 보였습니다. 단, Hits@10 지표에서는 실패했습니다.

LLM은  $N_i$ 라운드의 엔티티 예측을 수행합니다.

<sup>1</sup> <https://huggingface.co/meta-llama/Llama-2-7b-chat-hf>

모델	EN	FR	IT	JA	ZH	AVG
H 1	TransE	8.52	9.07	9.36	8.00	11.77
	유추	13.40	15.81	14.58	15.11	6.43
	ComplEx	10.92	11.75	11.49	13.95	17.79
	Distmult	6.89	7.73	7.93	8.16	5.66
	회전	24.08	24.61	25.57	29.49	31.36
	HAKE	31.64	32.92	30.99	35.53	52.24
	ICL	1.79	1.07	1.26	1.93	2.27
	GC-PLM	33.37	32.51	30.38	36.65	49.13
	DIFT (상급)	36.05	35.75	34.22	38.31	56.65
	우리	<b>36.50</b>	<b>36.72</b>	<b>35.93</b>	<b>41.60</b>	<b>58.63</b>
H 3	TransE	37.02	39.17	37.57	44.59	60.51
	유추	28.39	30.45	29.56	35.84	19.61
	ComplEx	23.05	23.92	23.05	29.68	41.12
	Distmult	14.59	14.82	15.79	19.16	17.94
	회전	40.73	42.13	41.68	49.78	62.36
	HAKE	43.30	43.27	41.88	47.52	63.22
	ICL	34.99	37.09	35.33	42.84	59.15
	GC-PLM	40.99	42.21	40.47	50.95	63.67
	DIFT (단일)	42.21	42.28	40.50	47.83	64.50
	우리	<b>46.25</b>	<b>45.30</b>	<b>44.22</b>	<b>51.97</b>	<b>66.93</b>
H 10	TransE	50.25	51.23	49.60	58.10	71.80
	유추	39.17	41.73	40.30	48.78	64.45
	ComplEx	34.84	37.51	35.70	45.02	59.97
	Distmult	26.74	26.78	27.41	36.04	50.25
	회전	52.66	53.17	51.50	<b>61.68</b>	<b>74.58</b>
	HAKE	53.37	52.06	50.49	57.85	70.04
	ICL	49.99	51.05	49.30	57.99	71.71
	GC-PLM	52.76	52.81	51.76	59.53	71.98
	DIFT (단일)	52.48	52.35	50.30	58.74	72.08
	우리	<b>54.71</b>	<b>53.45</b>	<b>52.31</b>	60.85	72.56
M R	TransE	24.85	25.97	25.46	28.38	37.51
	유추	22.85	25.01	23.82	27.44	21.04
	ComplEx	19.18	20.29	19.56	24.30	32.12
	Distmult	13.33	13.80	14.28	16.86	16.69
	회전	34.63	35.39	35.49	41.75	48.74
	HAKE	39.41	39.80	38.10	43.36	59.04
	ICL	20.11	20.70	19.80	23.98	31.06
	GC-PLM	36.66	37.21	37.18	42.21	55.39
	DIFT (단일)	40.99	40.64	39.09	44.54	61.38
	우리	<b>42.96</b>	<b>42.58</b>	<b>41.69</b>	<b>48.33</b>	<b>63.74</b>

표 2: 본 표는 다섯 가지 언어에 걸친 MKGC 결과를 제시한다. 임베딩 기반 방법인 TransE(Bordes et al., 2013), Analogy(Liu et al., 2017), ComplEx(Trouillon et al., 2016), DistMult(Yang et al., 2014), RotatE(Sun et al., 2019)는 모두 OpenKE 프레임워크(Han et al., 2018)를 사용하여 구현되었습니다. HAKE(Zhang et al., 2020)의 결과는 오픈소스 코드를 사용하여 재현되었습니다. ICL은 LLaMA-2-7b-chat 모델을 미세 조정 없이 사용한 평가를 의미합니다. GC-PLM(Song et al., 2023)은 현재 MKGC 분야의 최첨단(SOTA) 기법을 대표합니다. DIFT(Liu et al., 2024)는 LLM 기반 단일 언어 KGC 분야의 최첨단 기법입니다. ‘단일 언어(Single)’는 각 언어별로 독립적으로 모델을 훈련하는 방식을 의미합니다. 굵은 글씨로 표시된 숫자는 고려된 방법론과 언어 조합 중 최상의 결과를 나타냅니다.

Hits@10에서 RotatE의 성능을 뛰어넘었는데, 이는 주로 후보 엔티티 생성을 위해 상대적으로 성능이 낮은

모델	EN	FR	IT	JA	ZH	AVG
H	LoRAMoE	36.28	36.36	35.81	40.22	56.87
@	HydraLoRA	35.68	35.60	35.05	40.49	57.92
1	Ours	<b>36.50</b>	<b>36.72</b>	<b>35.93</b>	<b>41.60</b>	<b>58.63</b>
H	LoRAMoE	42.54	42.74	41.40	48.75	64.96
@	HydraLoRA	42.49	42.58	41.22	48.68	64.99
3	우리	<b>42.87</b>	<b>43.08</b>	<b>41.70</b>	<b>48.94</b>	<b>65.06</b>
H	LoRAMoE	52.27	52.50	50.91	59.07	72.19
@	HydraLoRA	52.51	52.41	50.81	<b>59.10</b>	72.24
10	Ours	<b>52.63</b>	<b>52.62</b>	<b>51.17</b>	59.09	<b>72.32</b>
M	LoRAMoE	41.15	41.13	40.30	45.90	61.52
R	HydraLoRA	40.80	40.62	39.81	46.06	62.18
R	Ours	<b>41.42</b>	<b>41.44</b>	<b>40.51</b>	<b>46.80</b>	<b>62.67</b>

표 3: 이 표는 KL-GMoE를 기준의 SOTA 미세 조정 방법인 LoRAMoE (Dou et al., 2024) 및 HydraLoRA (Tian et al., 2024)와 비교한 것입니다.

모델	훈련 가능한 매개변수 활성화된 매개변수 Lora 순위
TransE	106.1m
DIFT(LoRA)	159.9*5m
LoRAMoE	19.2m
HydraLoRA	12.5m
우리	32.9m
	9.4m
	4

표 4: 본 방법과 다른 방법들의 매개변수 수 비교를 보여줍니다.

후보 엔티티 생성에 상대적으로 성능이 낮은 TransE 모델을 사용했기 때문입니다. 후보 엔티티 검색 단계에서 TransE를 RotatE로 대체하고 실험을 수행했습니다. 해당 결과와 분석은 부록 A.3에 제시되어 있습니다. 기존 SOTA MKGC 방법인 GC-PLM과 비교했을 때, 우리의 프레임워크는 Hits@1, Hits@3, Hits@10 및 MRR 지표에서 각각 5.47%, 3.27%, 1.01%, 6.13%의 개선을 통해 상당한 성능 우위를 달성했습니다. 또한, 당사의 프레임워크는 SOTA LLM 기반 단일 언어 KGC 방법인 DIFT에 비해 성능이 크게 향상되었습니다. 전반적으로, 실험 결과는 당사가 제안한 프레임워크의 효과성과 우수성을 명확하게 입증합니다.

### 4.3 모델 아키텍처 비교 및 매개변수 분석

KL-GMoE 아키텍처와 기존 최첨단(SOTA) 미세 조정 방법(LoRAMoE 및 HydraLoRA 포함) 간의 MKGC 성능을 비교했습니다. 이 두 SOTA 방법은 모두 단일 쿼리를 처리하기 위해 다중 채널을 활용하는데, 이는 지식 분절화 문제를 초래할 수 있습니다. 표 3에서 볼 수 있듯이, KL-GMoE는

모델	평균 토큰 수	TFLOPs
LoRAMoE	353.89	2.37814
HydraLoRA	353.89	2.37580
DIFT(LoRA)	353.89	2.42721
Ours	353.89	2.37472

표 5: 본 방법과 다른 방법들의 계산 효율성을 비교한 표입니다. 제시된 데이터는 1,000개의 샘플로부터 계산된 평균값입니다.

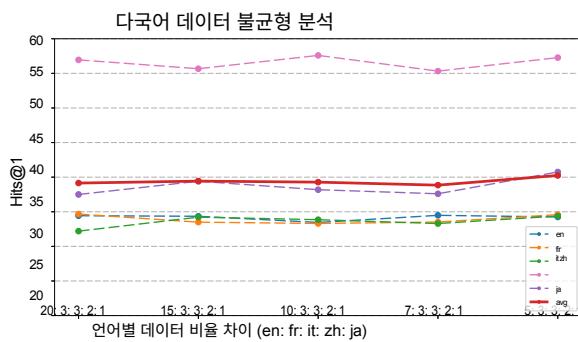


그림 3: 본 프레임워크의 Hits@1 점수 변화(훈련 데이터 설정별)

다섯 가지 다른 언어 비율로 진행되었습니다.

이러한 방법들은 평균 지표에 대해 수행되었습니다. 이 실험 결과는 우리의 방법이 지식 분열을 효과적으로 해결함으로써 MKGC 작업에서 성능을 향상시킨다는 것을 보여줍니다.

모델 매개변수 측면에서 KL-GMoE의 장점을 추가로 분석했습니다. 표 4에서 볼 수 있듯이, 임베딩 기반 방법인 TransE와 비교했을 때 KL-GMoE는 훈련 가능한 매개변수가 3.2배 적고 활성화된 매개변수가 11.3배 적습니다. LLM 기반 방법들 중 KL-GMoE는 다른 모든 방법보다 활성화 매개변수가 현저히 적습니다. 특히 DIFT와 비교했을 때 KL-GMoE는 훈련 가능 매개변수가 약 24.3배 적고 활성화 매개변수는 약 85.1배 적어 매개변수 수 측면에서 상당한 장점을 입증합니다. 동시에, 우리는 제안된 방법을 추론 중 FLOPs 측면에서 다른 방법들과 비교했습니다. 표 5에서 볼 수 있듯이, 우리의 방법은 현재 최첨단 LLM 기반 KGC 방법인 DIFT에 비해 FLOPs를 약 0.053 TFLOPs 정도 줄여, 우수한 계산 효율성을 입증합니다.

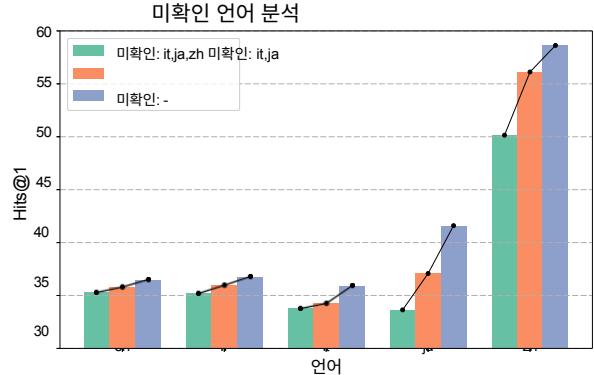


그림 4: 이 그림은 세 가지 다른 훈련 언어 설정에서 다섯 가지 언어에 대한 우리 방법의 Hits@1 성능을 보여줍니다.

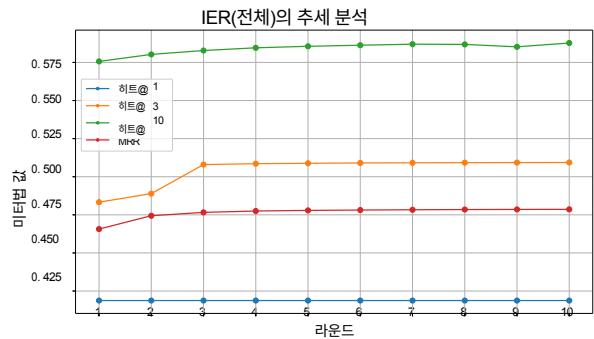


그림 5: 이 그림은 IER 방법의 반복 횟수가 성능에 미치는 영향을 보여줍니다.

#### 4.4 언어 불균형 분석

훈련 데이터에서 언어 분포가 불균형한 시나리오에서 프레임워크의 견고성을 평가하기 위해 실험을 수행했습니다. 구체적으로, 총 훈련 데이터 양은 일정하게 유지한 상태에서 다섯 가지 언어에 걸쳐 불균형한 훈련 데이터 비율을 적용한 실험을 수행했습니다. 그림 3에서 볼 수 있듯이, 언어 비율의 변화에도 불구하고 각 언어별 Hits@1 점수(점선)와 다섯 언어 평균 점수(빨간색 실선)는 비교적 안정적으로 유지되었습니다. 이는 본 프레임워크가 언어 분포 변동에 민감하지 않음을 보여줍니다. 이러한 분석을 바탕으로 본 프레임워크는 언어 간 공유 지식을 효과적으로 활용하여 강력한 견고성을 입증합니다.

#### 4.5 미검증 언어 분석

제안된 프레임워크의 훈련 데이터에 포함되지 않은 언어에 대한 일반화 능력을 평가하기 위해 분석 실험을 수행했습니다. 이 실험은 세 가지 서로 다른 훈련 구성으로 진행되었습니다: (1) 훈련된

언어		쿼리		공유	답변	DIFT 예측	우리의 예측
EN	(타월헤드, 작곡가, ?)	FR	Thomas 뉴먼	데이터드 키타이   타월헤드   토마스 뉴먼   제임스 뉴턴 하워드   크리스 메시나   롬프 캔트   캐리 프레스턴		<b>토마스 뉴먼</b>   롬프 캔트   제임스 뉴턴 하워드   マイ클 다나   시어도어 샤피로   마크 이삼   카터 베렐	
FR	(테라야마 슈지, 직업, ?)	IT	작사가	연출가   기자   <b>작사가</b>   영화 제작자   음악가   작곡가   싱어송라이터		<b>작사가</b>   영화 제작자   기타리스트   연출가   작곡가   기자   아티스트	
IT	(픽사 스토리, 연주자, ?)	EN	존 레스터	조 랜프트   에릭 라슨   <b>존 라세터</b>   밀트 칼   글렌 키언   제임스 알가   돈 한		<b>존 라세터</b>   에릭 라슨   밀트 칼   글렌 키언   크리스 벡   콜린 행크스   제임스 알가	
JA	(파울루스, 동일과 되는事物, ?)	ZH	파울루스	파울루스   파울로스   <b>파울</b>   폴라   크리스틴   콘스탄틴   크리스티나		<b>파울</b>   파울루스   폴라   파울로스   크리스틴 콘스탄틴   크리스티나	
ZH	(라코시·마가시, 구어, 문어, ?)	JA	러시아어	헝가리   라코시 마가시   <b>러시아어</b>   독일어   중국어   스페인어   라틴어		<b>러시아어</b>   독일어   일본어   영어   라틴어   헝가리어   폴란드어	

그림 6: 본 그림은 지식 공유 사례에서 본 방법과 DIFT의 예측 결과를 비교한 것이다. '공유' 열은 해당 쿼리의 지식이 LLM 훈련 데이터에 존재하지만 다른 언어로 제시되었음을 나타낸다.

(1) EN과 FR로 훈련된 모델; (2) EN, FR, ZH로 훈련된 모델;  
그리고  
(3) 다섯 가지 언어로 훈련됨. 그림 4에서 보듯이, 녹색 막대는  
영어(EN)와 프랑스어(FR) 데이터만으로 훈련된 대규모 언어  
모델(LLM)이 미학습 언어인 이탈리아어(IT), 일본어(JA), 중  
국어(ZH)에서 상당한 지식 그래프 연결(KGC) 성능을 보였음  
을 나타냅니다. 이는 지식 공유가 LLM 훈련 중 접한 언어 간  
뿐만 아니라 미학습 언어 간에도 상당한 교차 언어 일반화 능  
력을 효과적으로 발휘함을 명확히 보여줍니다. 또한 훈련 언  
어 수가 증가함에 따라 모든 언어에서 성능이 일관되게 향상  
되는 것을 관찰했습니다. 이 결과는 더 많은 언어로 된 훈련  
데이터가 LLM에 풍부한 지식 신호를 제공하여 다국어 지식  
공유를 촉진함을 시사합니다.

#### 4.6 IER 동향 분석

IER 방법에서 반복 횟수가 MKGC 성능에 미치는 영향을 평  
가하기 위해 분석적 실험을 수행하였다. 그림 5는 반복 횟수  
증가에 따른 모든 지표의 변화를 보여준다. 결과에서 알 수 있  
듯이, Hits@3, Hits@10 및 MRR은 처음 세 번의 반복에서 크  
게 개선되었으며 열 번째 반복에서 최적값에 도달했습니다.  
이러한 추세는 반복 횟수가 증가함에 따라 IER이 LLM이 다  
국어 공유 지식을 보다 효과적으로 활용하도록 하여 MKGC  
의 성능을 크게 향상시킨다는 것을 시사합니다.

#### 4.7 제거 실험

제안된 프레임워크 내 각 구성 요소의 효과성을 검증하기 위  
해 제거 실험을 수행했습니다. 각 구성 요소를 순차적으로 제  
거하여 그 기여도를 평가했습니다. 표 6에서 볼 수 있듯이,  
KL-GMoE 구성 요소를 제거하면 Hits@1이 41.88에서  
40.28로, Hits@3이 50.93에서 49.71로, Hits@10이 58.78에  
서

모델	H@1	H@3	H@10	MRR
본 연구	<b>41.88</b>	<b>50.93</b>	<b>58.78</b>	<b>47.86</b>
kg 제외 우리 모델	40.28	49.71	58.07	46.55
kg+ier 제외	40.28	47.66	57.29	45.42

표 6: 본 표는 KL-GMoE(**kg**) 및 IER(**ier**) 구성요소에 대한 제거 실험  
결과를 보여줍니다. 모든 결과는 다섯 가지 언어 지표의 평균값입니  
다.

58.07로, MRR은 47.86에서 46.55로 감소했습니다. 이는  
KL-GMoE 구성 요소가 MKGC 성능 향상에 매우 중요함을  
나타냅니다. 또한 KL-GMoE와 IER을 동시에 제거했을 때,  
KL-GMoE만 제거했을 때보다 Hits@3, Hits@10, MRR 값이  
더 크게 감소했습니다. 이는 IER 구성 요소 역시 MKGC 성  
능 향상에 긍정적인 기여를 함을 보여줍니다. 이러한 제거 실  
험 결과는 제안된 KL-GMoE 및 IER 구성 요소의 효과성을  
강력히 입증합니다.

#### 4.8 사례 연구

우리는 다국어 지식 공유에서 프레임워크의 성능을 평가하기  
위한 사례 연구를 수행했습니다. 이 사례의 질의는 LLM이  
훈련 과정에서 학습한 지식이지만 다른 언어로 표현된 것입  
니다. 그림 6에서 볼 수 있듯이, 영어 질의(*Towelhead,  
composer, ?*)에 대해 LLM은 이미 프랑스어 훈련 데이터에서  
이 지식을 학습했습니다. 우리 프레임워크는 이 프랑스어 지  
식을 성공적으로 활용하여 해당 엔티티를 **토마스 뉴먼**으로  
정확히 예측했습니다. 반면, 최첨단 LLM 기반 방법들은 **데  
이비드 키터이**로 잘못 예측했습니다. 이는 우리 프레임워크  
가 교차언어 공유 지식을 효과적으로 활용하여 완성 정확도  
를 향상시킬 수 있음을 보여줍니다.

## 5 관련 연구

임베딩 기반 방법은 지식 그래프(KG) 내 엔티티와 관계를 저차원 벡터 공간으로 매핑합니다. 예를 들어, TransE(Bordes et al., 2013)는 엔티티와 관계의 번역 원리에 기반합니다. RotatE(Sun et al., 2019)는 각 관계를 복소 벡터 공간에서의 회전으로 처리합니다. DMoG(Song et al., 2022a)는 온톨로지와 텍스트 그래프를 융합하여 사실 그래프의 미관측 관계를 표현합니다. TransH(Wang et al., 2014)는 관계를 초평면으로 모델링합니다. HOLEX(Xue et al., 2018)는 높은 모델복잡도 방법과 HolE(Nickel et al., 2016) 사이를 보간합니다. TR-GCN(Song et al., 2022b)은 관찰되지 않은 관계를 표현하기 위한 온톨로지 기반 제로샷 관계 학습 방법을 제안합니다.

생성 기반 방법은 지식 그래프 연결 예측(KGC) 작업을 텍스트 생성 작업으로 변환합니다. 예를 들어, KGT5(Sax-ena 외, 2022)는 KG 링크 예측을 시퀀스-투-시퀀스 작업으로 제시합니다. GC-PLM(Song et al., 2023)은 전역 및 지역 지식 제약 조건을 도입하여 MKGC의 성능을 향상시킵니다. GenKGC(Xie et al., 2022)는 관계 기반 증명과 엔티티 인식의 계층적 디코딩 전략을 도입합니다. KICGPT (Wei et al., 2023)는 LLM과 KGE 모델을 통합하고, 지식 기반 문맥 학습 전략을 채택하여 여러 엔티티의 순위를 재조정합니다. DIFT (Liu et al., 2024)는 LoRA (Hu et al., 2022)를 사용하여 LLM을 미세 조정하여 KGE 모델에서 얻은 후보 엔티티 중에서 가장 최적의 엔티티를 선택합니다.

## 6 결론

본 논문에서는 새로운 MKGC 프레임워크를 제안한다. 이 프레임워크는 KL-GMoE와 IER라는 두 가지 구성 요소를 통합한다. KL-GMoE는 언어 간 공유 지식을 효율적으로 포착함으로써 완성 성능을 크게 향상시킨다. IER는 다중 반복적 접근법을 통해 교차 언어 공유 지식을 완전히 활용하여 완성 성능을 더욱 개선한다. 실험 결과는 본 프레임워크가 MKGC 작업에서 우수한 성능을 보임을 입증한다.

## 제한 사항

본 프레임워크는 LLM의 토큰 길이에 제한을 받기 때문에 KG 내 모든 엔티티를 기반으로 엔티티 선택을 수행할 수 없습니다. 또한 본 프레임워크는 텍스트 정보만을 처리합니다. 이 한계로 인해 다른 모달리티의 정보를 통합할 수 없어

다중 모달 KG 데이터셋에 적용하는 데 방해가 됩니다.

## 윤리 성명

본 논문은 MKGC를 위한 방법을 제안하고 공개 데이터셋에서 확장된 다국어 데이터셋에 대한 실험을 수행합니다. 따라서 이 시나리오에서는 데이터 프라이버시 문제가 존재하지 않습니다.

## 감사의 말

본 연구는 중국 국가자연과학기금회( U21B2027, U23A2038, 62166023, 62376270), 원난성 주요 과학기술 특별계획 프로젝트( 202402AG050007, 202502AD080012, 202502AD080016), 원난성 기초연구 일반 프로젝트(지원 번호: 202301AS070047, 202201BE070001-021).

## 참고문헌

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. *Translating embeddings for modeling multi-relational data*. *Advances in neural information processing systems*, 26.

천천, 왕위페이, 리빙, 램궈옌. 2022. *지식은 평평하다: 다양한 지식 그래프 완성을 위한 Seq2Seq 생성 프레임워크*. 제29회 국제계산언어학회 학술대회 논문집, 4005–4017쪽, 대한민국 경주. 국제계산언어학회.

두 시한, 저우 엔위, 류 얀, 가오 쟁양, 쉘 웨이, 승 리마오, 저우 유하오, 왕 샤오, 시 즈형, 판 샤오란, 푸 시량, 주 장, 정 루이, 구이 타오, 장 치, 황 쑤안징. 2024. *LoRAMoE: 대규모 언어 모델에서 세계 지식 망각 완화를 위한 MoE 스타일 플러그인*. 제62회 전산언어학회 연차대회 논문집(제1권: 장문 논문), 1932–1945쪽, 태국 방콕. 전산언어학회.

Xiou Ge, Yun Cheng Wang, Bin Wang, C-C Jay Kuo 외 1명. 2024. *지식 그래프 임베딩: 개요*. *APSIPA Transactions on Signal and Information Processing*, 13(1).

Xu Han, Shulin Cao, Lv Xin, Yankai Lin, Zhiyuan Liu, Maosong Sun, Juanzi Li. 2018. *OpenKE: 지식 임베딩을 위한 오픈 툴킷*. EMNLP 회의록.

- 한재위, 가오차오, 류진양, 장제프, 장사이чен. 2024. **대규모 모델을 위한 매개변수 효율적 미세 조정: 포괄적 서베이**. *arXiv 사전 인쇄본 arXiv:2403.14608*.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen. 2022. **LoRA: 대규모 언어 모델의 저순위 적응**. *국제 표현 학습 컨퍼런스*.
- 평후, 시즈 리우, 창장 가오, 신 황, 슈에 한, 준란 평, 차오 덩, 슈지안 황. 2025. **대규모 언어 모델은 언어 간 지식 없는 추론자이다. 2025년 전미계산언어학회 아메리카 국가별 지부 학술대회: 인간 언어 기술 (제1권: 장문 논문집)**, 1525–1542쪽, 뉴멕시코주 앨버커키. 전미계산언어학회.
- 황카이위, 모평란, 장신위, 리홍량, 리유, 장위안치, 이웨이젠, 마오위 룽, 류진천, 쉬위좡 외 1인. 2024. **다국어 기능을 갖춘 대규모 언어 모델에 관한 연구: 최근 발전과 새로운 지평**. *arXiv 사전 인쇄본 arXiv:2405.10936*.
- 지샤오슝, 판시루이, 에릭 캄브리아, 페카 마르티넨, 필립 S. 유. 2022. **지식 그래프에 관한 연구: 표현, 획득 및 응용**. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494–514.
- 옌스 레만, 로버트 이젤, 막스 야콥, 안야 엔츠슈, 디미트리스 콘토코 스타스, 파블로 엔 멘데스, 세바스티안 헬만, 모하메드 모르시, 패트릭 반 클리프, 쇠렌 아우어 외 1인. 2015. **Dbpedia–위키피디아에서 추출한 대규모 다국어 지식베이스**. *세미黠 웹*, 6(2):167–195.
- 한사오 리우, 위에신 우, 이밍 양. 2017. **다중 관계 임베딩을 위한 아날로그적 추론**. *제34회 국제 기계 학습 컨퍼런스 논문집 - 제70권, ICML'17*, 2168–2178쪽. JMLR.org.
- 양 리우, 샤오빈 티안, 제쿤 쑨, 웨이 후. 2024. **지식 그래프 완성을 위한 판별 지침을 활용한 생성형 대규모 언어 모델의 미세 조정**. *제 시맨틱 웹 컨퍼런스*, 199–217쪽. Springer.
- 막시밀리안 니켈, 로렌초 로사스코, 토마소 포지오. 2016. **지식 그래프의 흘로그래픽 임베딩: 인공 지능에 관한 AAAI 학회 논문집**, 제 30권.
- Apoorv Saxena, Adrian Kochsiek, and Rainer Gemulla. 2022. **시퀀스-투-시퀀스 지식 그래프 완성 및 질문 답변**. *제60회 전산언어학회 연차대회 논문집 (제1권: 장문 논문)*, 2814–2828쪽, 아일랜드 더블린. 전산언어학회.
- 송란, 가오성샹, 가오샤오판이, 마오쿤리, 위정타오. 2025. **Mike-pLLM: 사전 훈련된 대규모 언어 모델을 위한 다국어 지식 편집 벤치마크**. *신경계산*, 651:130979.
- Ran Song, Shizhu He, Shengxiang Gao, Li Cai, Kang Liu, Zhengtao Yu, Jun Zhao. 2023. **지식 제약 조건을 적용한 사전 훈련된 언어 모델로부터의 다국어 지식 그래프 완성**. *컴퓨터 언어학 협회 연구 결과: ACL 2023*, 7709–7721쪽, 캐나다 토론토. 컴퓨터 언어학 협회.
- 송란, 허시주, 정순콩, 가오성샹, 류강, 위정타오, 조준. 2022a. **그래프 훈합 모델의 분리: 온톨로지 및 텍스트 전문가 융합을 통한 지식 그래프 완성을 위한 미관측 관계 학습**. *제29회 국제계산언어학회 (ICL) 논문집*, 2237–2246쪽, 대한민국 경주. 국제계산언어학회 (ICL).
- 송란, 허시주, 정순콩, 가오성샹, 류강, 자오준, 위정타오. 2022b. **지식 그래프 제로샷 관계 학습을 위한 온톨로지 기반 및 텍스트 강화 표현**. *ICLR 2022 자연어 처리용 그래프 기반 딥러닝 워크숍*.
- Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. **Rotate: 복잡한 공간에서의 관계 회전을 통한 지식 그래프 임베딩**. *국제 표현 학습 컨퍼런스*.
- 천춘린, 시잔, 구오즈장, 리리, 쉬청종. 2024. **Hydralora: 효율적인 미세 조정을 위한 비대칭 LORA 아키텍처**. *신경 정보 처리 시스템 발전 (NeurIPS)*.
- 테오 트루이옹, 요하네스 벨블, 세바스티안 리델, 에릭 고시에, 기욤 부샤르. 2016. **단순한 링크 예측을 위한 복잡한 임베딩**. *제33회 국제 기계 학습 학회(ICML) 논문집 - 제48권, ICML'16*, 2071–2080쪽. JMLR.org.
- Denny Vrandecic' and Markus Krötzsch. 2014. **위키데이터: 자유 협업 지식 기반**. *ACM 커뮤니케이션즈*, 57(10):78–85.
- 왕샤오즈, 가오텐위, 주차오청, 장정얀, 류즈위안, 리쥔즈, 탕젠. 2021. **KEPLER: 지식 임베딩과 사전 훈련된 언어 표현을 위한 통합 모델**. *계산언어학회지*, 9:176–194.
- 왕전, 장젠원, 평젠틴, 천정. 2014. **초평면 상의 변환을 통한 지식 그래프 임베딩**. *제28회 AAAI 인공 지능 학회 논문집*, AAAI'14, 1112–1119쪽. AAAI Press.

Yanbin Wei, Qiushi Huang, Yu Zhang, and James Kwok. 2023. KICGPT: 지식 그래프 완성을 위한 문맥 내 지식을 갖춘 대규모 언어 모델. *계산언어학회 논문집: EMNLP 2023*, 8667–8683쪽, 싱가포르. 계산언어학회.

Gerhard Weikum. 2021. *지식 그래프 2021: 데이터의 오디세이*. *Proceedings of the VLDB Endowment*, 14(12):3233–3238.

Xin Xie, Ningyu Zhang, Zhoubo Li, Shumin Deng, Hui Chen, Feiyu Xiong, Mosha Chen, and Huajun Chen. 2022. 차별화에서 생성으로: 생성형 트랜스포머를 활용한 지식 그래프 완성. *웹 컨퍼런스 2022 동반 논문집*, WWW '22, 162–165쪽, 뉴욕, NY. 미국. 컴퓨터 학회.

Yexiang Xue, Yang Yuan, Zhitian Xu, Ashish Sabharwal. 2018. 지식 완성을 위한 홀로그래픽 임베딩 확장. *신경정보처리시스템 발전*, 31.

양비산, 이원타우, 허샤오동, 가오젠펑, 덩리. 2014. 지식 기반 학습 및 추론을 위한 엔티티 및 관계 임베딩. *국제 표현 학습 컨퍼런스*.

장잔치우, 차이진위, 장용동, 왕지에. 2020. 링크 예측을 위한 계층 인식 지식 그래프 임베딩 학습. *인공 지능에 관한 AAAI 컨퍼런스 논문집*, 제34권, 3065–3072쪽.

Wenxuan Zhou, Fangyu Liu, Ivan Vulic', Nigel Collier, and Muhamao Chen. 2022. Prix-LM: 다국어 지식 기반 구축을 위한 사전 훈련. *제60회 전산언어학회 연차총회 논문집(제1권: 장문 논문)*, 5412–5424쪽, 아일랜드 더블린. 전산언어학회.

## A 부록

### A.1 프롬프트 예시

훈련 단계에서 다양한 길이의 후보 엔티티 목록에 대한 프롬프트 예시를 제시합니다.

---

프롬프트     꼬리 엔티티 t가 누락된 삼중항을 주어짐: (Saint George and the Dragon, material used, t).

다음은 실체 '성 조지와 용'에 대한 설명 정보입니다:

성 조지와 용, 성 조지와 용 또는 용을 죽이는 성 조지는 베네치아 화가 틴토레토가 1555년 또는 1558년에 그린 그림이다. 이후 영국 수집가가 이를 구입했다.

성 조지와 용을 묘사한 삼중주 작품들:

[(성 조지와 용, 묘사, 언덕); (성 조지와 용, 묘사, 창); (성 조지와 용, 창작자, 야코포 틴토레토); (성 조지와 용, 묘사, 전투); (성 조지와 용, 묘사, 여성); (성 조지와 용, 묘사, 하늘)]

t의 실체명은 무엇인가요? 아래 실체 목록에서 하나를 선택하세요: [유화; 성 조지와 용; 나무; 템페라; 직물; 벽돌; 진주; 변성암; 편암; 사암; 페인트; 화성암; 조직; 보석; 활동; 구리; 직조 직물; 화산암; 대리석; 용; 현무암; 퇴적암; 세 여신; 석회암; 강철]

[답변]:

---

번호     25

개 엔티티

---

프롬프트     꼬리 엔티티 t가 누락된 삼중항 (Jason Lee, instance of, t)이 주어집니다. 다음은 엔티티 Jason Lee에 대한 설명 정보를 제공합니다:

제이슨 리(Jason Lee, Jason Michael Lee, 1970년 4월 25일 출생)는 미국의 배우, 사진작가, 프로듀서, 스케이트보더, 코미디언, 작가이다. 그는 TV 시리즈 <아메리칸 드림즈>에서 얼 히키 역으로 가장 잘 알려져 있다.

다음은 엔티티 제이슨 리를 포함하는 삼중항목입니다:

[(몰래츠 출연진, 제이슨 리); (제이슨 리, 민족, 스코틀랜드계 미국인); (제이슨 리, 직업, 시나리오 작가); (제이슨 리, 직업, 배우); (제이슨 리, 직업, 영화 제작자); (제이슨 리, 직업, 사업가)]

t의 엔티티 이름은 무엇입니까? 아래 엔티티 목록에서 하나를 선택하세요: [제이슨 리; 인간; 쌍둥이; 제이슨 알렉산더; 소피아 베르가라; 케빈 스미스; 스크린 액터스 길드 어워드; 데이비드 크로스; 3D 영화; 대학; 프라임타임 에미 어워드; 스포츠 시즌; MTV 무비 어워드; 케일리 쿠오코; 스페인 지방 자치 단체; 제이슨 뮤즈; 10년 단위; 군 계급; 교외 지역; 애니메이션 스튜디오; 제인 린치; 행크 아자리아; 여성 영화상; 브레킨 마이어; 내 이름은 얼; 패트릭 워버튼; 사업]

[답변]:

---

개체 수     27

개 항목

---

## A.2 반복적 엔티티 재정렬 알고리즘의 세부 사항

---

### 알고리즘 1 반복적 엔티티 재정렬(IER)

---

- 1: **입력:** 쿼리  $q = (h, r, ?)$ ,  $M^{(1)} = [e_1, e_2, \dots, e_m]$ : KGE 모델이 생성한 상위  $m$  개 엔티티,
  - $N, L^{(1)} = M^{(1)}$
  - 2:  **$t = 1$ 부터  $N_t$  까지 반복**
  - 3:  $e^{(t)} = \operatorname{argmax}_{e_i} P(e_i | h, r, M^{(t)})$ ;
  - 4:  $M^{((t+1))} = M^{((t))} \setminus \{e^{(t)}\}$ ;
  - 5:  $L^{(t+1)} = \operatorname{Insert}(L^{(t)} \setminus \{e^{(t)}\}, t, e^{(t)})$ ;
  - 6: **for 종료**
  - 7: **출력:**  $L^{(Nt+1)}$
- 

## A.3 지식 그래프 임베딩 모델 분석

표 8의 실험 결과는 RotatE를 사용하여 후보 엔티티를 검색할 때, 제안된 방법이 기존 RotatE 모델 대비 Hits@1에서 14.78% 증가하는 등 상당한 성능 향상을 달성했음을 명확히 보여줍니다. 특히 Ours+RotatE는 여러 언어별 지표에서 Ours+TransE보다 약간 낮은 성능을 보입니다. 이 현상은 각 KGE 모델이 생성한 후보 집합 내에서 올바른 엔티티의 상위 1위 순위 비율이 다르기 때문으로 볼 수 있습니다. 구체적으로, TransE를 사용할 때 1순위로 랭킹된 정답 엔티티의 비율은 14.38%였으나, RotatE를 사용하면 이 비율이 30.51%로 크게 증가했습니다. 따라서 이 특징은 어느 정도 영향을 미쳤습니다: 파인 튜닝 단계에서 LLM은 후보 집합에서 1순위로 랭킹된 엔티티를 최종 답안으로 선택하는 경향이 더 강해집니다. 우리는 이러한 "탑-1 편향"이 어느 정도 모델의 다른 잠재적 정답 탐색을 억제하여 일부 언어에서 Ours+RotatE가 Ours+TransE보다 약간 낮은 성능을 보일 수 있다고 가정합니다. 향후 연구에서는 기존 KGE 모델에 의존하지 않는 보다 안정적인 파인튜닝 지시어 세트 구축 방안을 추가로 탐구할 계획입니다.

모델		EN	FR	IT	JA	ZH	AVG
H	회전	24.08	24.61	25.57	29.49	31.36	27.02
@1	우리+TransE	36.50	<b>36.72</b>	<b>35.93</b>	41.60	58.63	<b>41.88</b>
	우리+회전	<b>36.55</b>	35.70	35.46	<b>41.78</b>	<b>59.49</b>	41.80
H	회전	40.73	42.13	41.68	49.78	62.36	47.34
@3	우리+TransE	46.25	<b>45.30</b>	<b>44.22</b>	51.97	66.93	50.93
	우리+회전	<b>46.38</b>	45.05	44.20	<b>52.27</b>	<b>67.13</b>	<b>51.01</b>
H	회전	52.66	53.17	51.50	61.68	74.58	58.72
@10	우리+전환	54.71	53.45	52.31	60.85	72.56	58.78
	우리+	<b>54.76</b>	<b>54.13</b>	<b>52.53</b>	<b>62.31</b>	<b>74.59</b>	<b>59.66</b>
	회전						
M	회전	34.63	35.39	35.49	41.75	48.74	39.20
R	우리+전환	42.96	<b>42.58</b>	<b>41.69</b>	48.33	63.74	47.86
R	우리+	<b>43.04</b>	42.18	41.58	<b>49.01</b>	<b>64.46</b>	<b>48.05</b>
	회전						

표 8: 후보 엔티티 검색 과정에서 제안된 프레임워크의 성능에 대한 다양한 KGE 모델의 영향.

## A.4 전문가 라우팅 분석

지식 공유의 존재를 검증하기 위해 테스트 샘플에서 전문가 선택을 분석했습니다. 각 전문가 라우팅의 출력을 획득하고 언어적 및 지식적 측면 모두에서 시각화했습니다.

차원. 그림 7의 왼쪽은 언어적 차원을 기반으로 한 전문가 라우팅 분석을 보여줍니다. 이는 서로 다른 언어의 샘플에 대해 LLM의 각 트랜스포머 레이어에서 전문가 선택 과정을 설명합니다. 다양한 언어 샘플에 걸친 전문가 선택은 대체로 일관성을 보이며, 이는 MKGC 작업에서 본 방법이 언어 간 차별을 두지 않음을 시사합니다. 그림 우측은 각 관계별 전문가 선택을 보여줍니다. 분석 결과, 서로 다른 언어에서 동일한 관계를 가진 샘플은 대체로 동일한 전문가들에 의해 처리되는 것으로 나타났습니다. 전반적으로 일부 관계는 일관된 전문가 선택을 보인 반면, 다른 관계들은 차이를 보였습니다. 이 분석을 바탕으로 본 방법이 서로 다른 언어 간 지식 공유를 효과적으로 활용할 수 있음을 검증합니다.

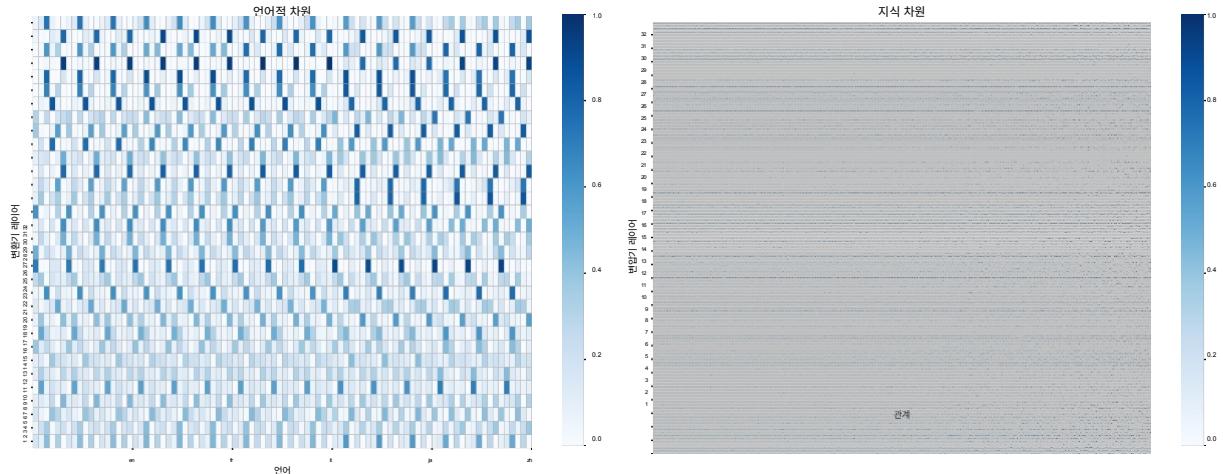


그림 7: 왼쪽은 다섯 가지 언어에 걸친 전문가 선택을 보여줍니다. 가로축은 모든 언어를 나타내며, 각 언어 내의 작은 막대는 전문가 한 명을 나타냅니다. 세로축은 LLM의 트랜스포머 내 레이어 번호를 나타냅니다. 각 블록의 색상 농도는 특정 전문가를 선택한 샘플의 빈도를 나타냅니다. 오른쪽은 모든 관계에 걸친 전문가 선택을 보여줍니다. 가로축은 모든 관계를 나타냅니다. 세로축은 트랜스포머 내 레이어 번호를 나타내며, 각 레이어 내 행은 전문가 한 명을 나타냅니다.

## A.5 IER 추세 분석

우리는 IER 방법에서 반복 횟수가 다섯 가지 언어 평가 지표에 미치는 영향을 분석했습니다. 실험 결과는 반복 횟수 증가에 따라 모든 언어에서 성능이 뚜렷이 향상됨을 보여줍니다. 이러한 결과는 IER 방법이 언어 간 공유 지식을 활용할 수 있는 능력을 입증합니다.

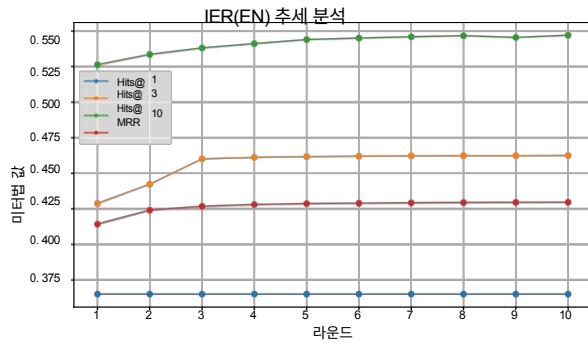


그림 8: 이 그림은 영어 테스트 세트에서 반복 횟수에 따른 IER 방법의 메트릭 성능 변화를 보여줍니다.

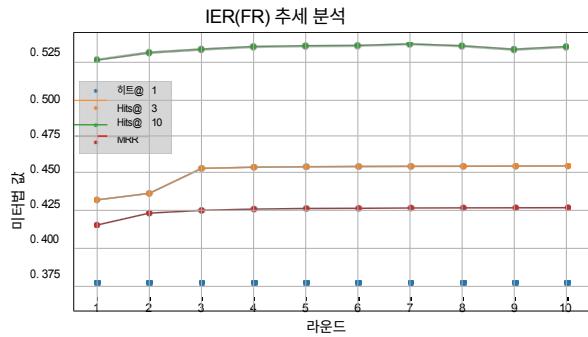


그림 9: 이 그림은 프랑스어 테스트 세트에서 반복 횟수에 따른 IER 방법의 메트릭 성능 변화를 보여줍니다.

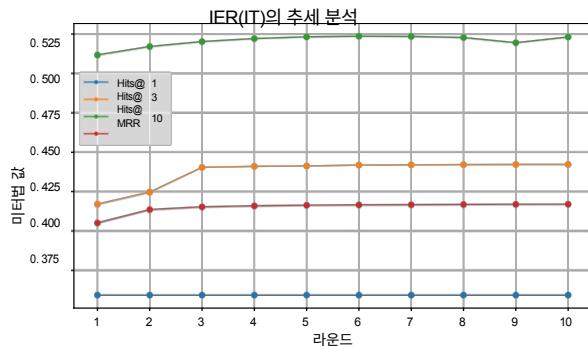


그림 10: 이 그림은 이탈리아어 테스트 세트에서 반복 횟수에 따른 IER 방법의 메트릭 성능 변화를 보여줍니다.

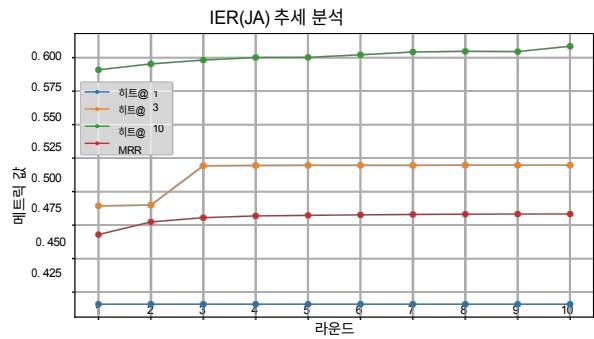


그림 11: 이 그림은 일본어 테스트 세트에서 반복 횟수에 따른 IER 방법의 메트릭 성능 변화를 보여줍니다.

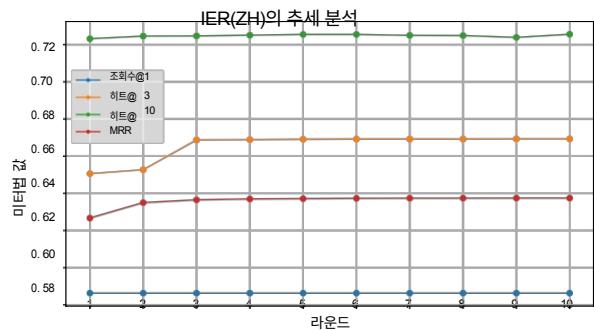


그림 12: 이 그림은 중국어 테스트 세트에서 IER 방법의 메트릭 성능이 반복 횟수에 따라 어떻게 변화하는지 보여줍니다.