



# Leveraging LLM-derived Knowledge Priors for Temporal Knowledge Graph Forecasting

시간적 지식 그래프 예측을 위한 LLM 기반 지식 사전 정보 활용

**Seyeong Kim (김세영)**

Graph & Language Intelligence Laboratory  
Department of Computer Science and Engineering  
Konkuk University

2026.07.22

# CONTENTS

## 1. Background

- Task Definition
- Dataset
- Metric

## 2. Approach

## 3. Why LLM?

## 4. LLM Integration Strategies

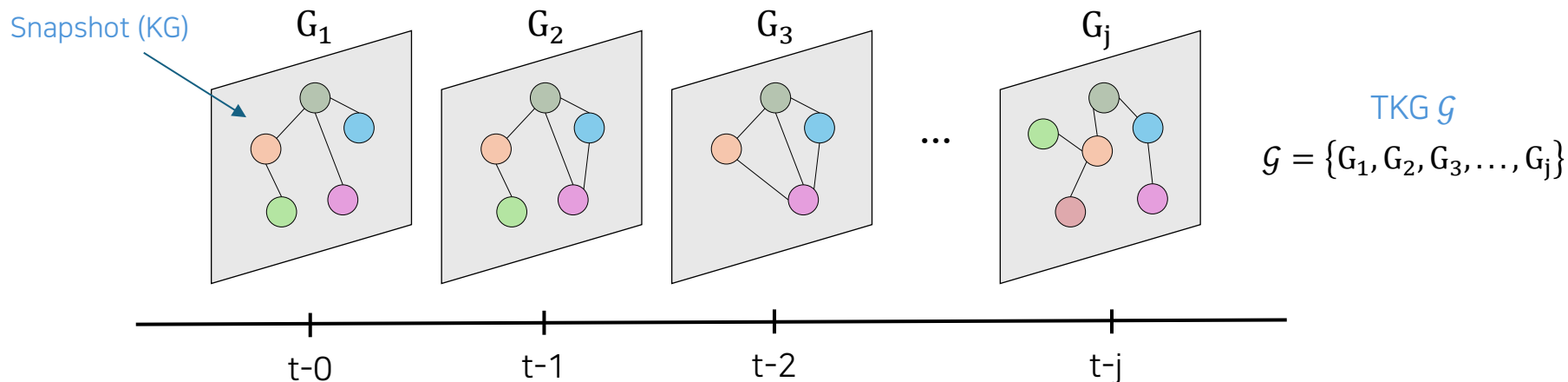
## 5. Research Direction



# Task Definition : What is Forecasting?

## Temporal Knowledge Graph

- Knowledge Graph
  - ✓ 현실 세계의 fact들을 (subject, relation, object)의 triple 형태로 표현
    - Ex) (Messi, plays\_for, Barcelona)
  - ✓ Fact들은 시간에 따라 변화하는데 KG는 이를 표현하지 못 함
    - Ex) (Messi, plays\_for, Barcelona) 는 2020년엔 참이고 2022년엔 거짓
- Temporal Knowledge Graph
  - ✓ 기존 triple에 timestamp를 붙여 (subject, relation, object, **time**)의 quadruple 형태로 표현
    - Ex) (Messi, plays\_for, Barcelona, **2020**) / (Messi, plays\_for, PSG, **2022**)



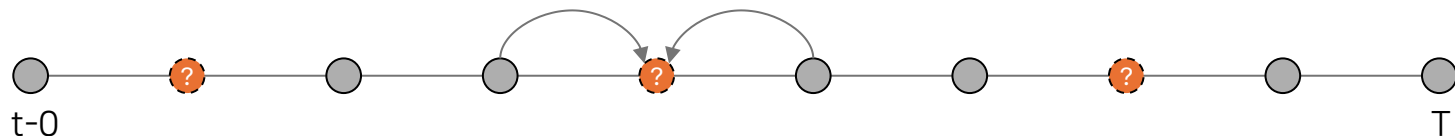
# Task Definition : What is Forecasting?

## Interpolation vs. Extrapolation

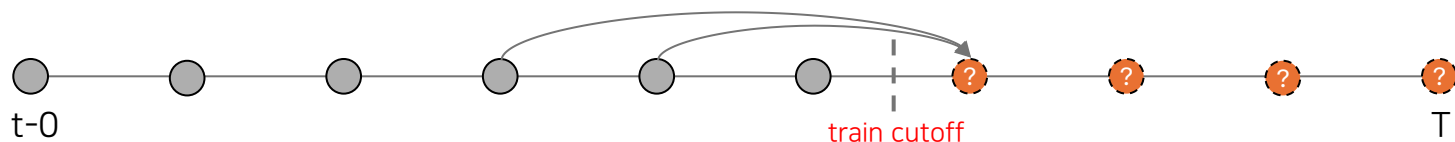
- Interpolation
  - ✓ 관측 구간 내 어떤 시점의 fact를 추론
    - 시간 순서를 랜덤으로 섞어 데이터를 split
    - 쿼리에 대해 양방향 (과거 + 미래)으로 참고 가능
- Extrapolation (Forecasting)
  - ✓ 관측 구간 이후 시점의 fact를 추론
    - 시간 순서대로 fact들을 나열하여 앞 쪽에서 부터 train / validation / test로 나눔
    - 쿼리에 대해 단방향 (과거)으로 참고 가능

● known fact    ● fact to predict

Interpolation



Extrapolation



# Task Definition : What is Forecasting?

## Query Type

- Entity Prediction

- ✓ 미래 시점  $t$ 의 fact에서 entity 하나를 ?로 두고 나머지를 조건으로 줌
  - $(?, \text{relation}, \text{object}, t)$
  - $(\text{subject}, \text{relation}, ?, t)$
- ✓ 각 fact마다 inverse quadruple을 데이터에 함께 추가  $\text{relation}^{-1} \text{ ID} = \text{relation ID} + |\text{relation}|$ 
  - $(\text{subject}, \text{relation}, \text{object}, t) \rightarrow (\text{object}, \text{relation}^{-1}, \text{subject}, t)$
  - 모델은 object 예측 한 가지만 처리하면 됨

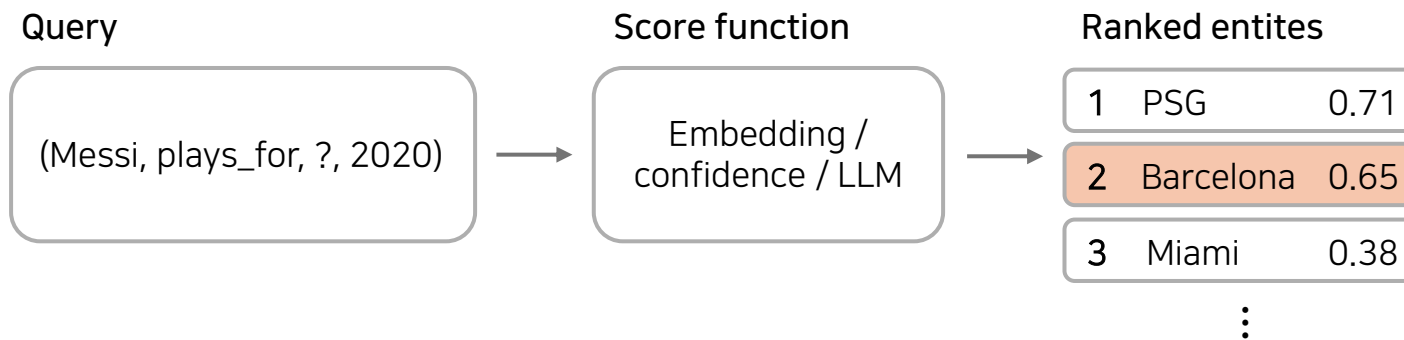
- Relation Prediction

- ✓ 미래 시점  $t$ 의 fact에서 relation 하나를 ?로 두고 나머지를 조건으로 줌
  - $(\text{subject}, ?, \text{object}, t)$
- ✓ 데이터 내 relation 수가 entity 수보다 훨씬 적어 변별력이 낮아 대부분의 논문에서 다루지 않음
  - ICEWS14 기준 relation이 230개
  - 소개할 논문들 역시 entity prediction을 기준으로 함

# Task Definition : What is Forecasting?

## Evaluation formulation

- Ranking
  - ✓ 모델은 정답 하나를 뽑는 게 아니라 모든 엔티티에 점수를 매김
    - $(s, r, ?, t) \rightarrow$  전체 엔티티에 대한 점수 벡터  $\rightarrow$  점수 순으로 정렬  $\rightarrow$  정답의 순위
- Score function
  - ✓ Embedding-based : 학습된 표현을 이용
  - ✓ Rule-based : rule confidence 집계
  - ✓ 이외에도 둘을 섞거나 LLM이 직접 개입하는 방법 존재



정답 Barcelona는 전체 엔티티 중 2위  $\rightarrow$  이 순위로 평가

# Dataset

## ICEWS (Integrated Crisis Early Warning System)

- 뉴스에서 추출한 전 세계 정치, 외교 사건
- 시간 단위 : 24시간
- 하루 = snapshot 하나
- 기간 별 데이터셋
  - ✓ ICEWS14 (2014년) / ICEWS05-15 (2005년 - 2015년) / ICEWS18 (2018년)

### Quadruple Example

(Xi Jinping, Make a visit, New Zealand, 2014-11-26)

$V$  : Entity set    $R$  : Relation set    $\mathcal{E}$  : Fact set

| Dataset    | $ V $  | $ R $ | $ \mathcal{E}_{train} $ | $ \mathcal{E}_{valid} $ | $ \mathcal{E}_{test} $ |
|------------|--------|-------|-------------------------|-------------------------|------------------------|
| ICEWS14    | 6,869  | 230   | 74,845                  | 8,514                   | 7,371                  |
| ICEWS05-15 | 10,094 | 251   | 368,868                 | 46,302                  | 46,159                 |
| ICEWS18    | 23,033 | 256   | 373,018                 | 45,995                  | 49,545                 |

# Dataset

## GDELT (Global Database of Events, Language, and Tone)

- 전 세계 방송, 인쇄, 웹 뉴스를 포괄하는 대규모 글로벌 이벤트
- 시간 단위 : 15분
- Entity는 적는데 fact는 압도적으로 많음
  - ✓ Entity 상당 수가 Police, Government 같은 추상 개념이라 예측이 근본적으로 어려움

### Quadruple Example

(Police, Consult, Government, 20180115143000)

$V$  : Entity set    $R$  : Relation set    $\mathcal{E}$  : Fact set

| Dataset | $ V $ | $ R $ | $ \mathcal{E}_{train} $ | $ \mathcal{E}_{valid} $ | $ \mathcal{E}_{test} $ |
|---------|-------|-------|-------------------------|-------------------------|------------------------|
| GDELT   | 7,691 | 240   | 1,734,399               | 238,765                 | 305,241                |

# Dataset

## WIKI / YAGO (WIKIdata / YAGO)

- 정치 사건이 아닌 일반 백과사전형 지식
- 시간 단위 : 1년
- 특정 event가 아니라 언제부터 언제까지 유효한지 state 형태
  - ✓ Snapshot 간 변화가 큼

### Quadruple Example

(Albert Einstein, worksAt, Princeton University, [1933,1955])

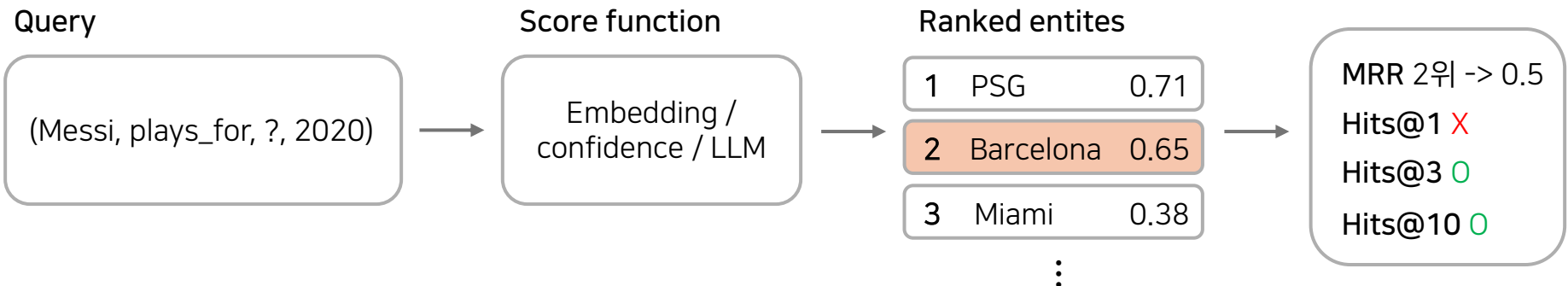
$V$  : Entity set    $R$  : Relation set    $\mathcal{E}$  : Fact set

| Dataset | $ V $  | $ R $ | $ \mathcal{E}_{train} $ | $ \mathcal{E}_{valid} $ | $ \mathcal{E}_{test} $ |
|---------|--------|-------|-------------------------|-------------------------|------------------------|
| WIKI    | 12,554 | 24    | 539,286                 | 67,538                  | 63,110                 |
| YAGO    | 10,623 | 10    | 161,540                 | 19,523                  | 20,026                 |

# Metric

## MRR / Hits@k

- MRR
  - ✓ 정답 순위의 역수를 모든 query에 대해 평균
  - ✓ 상위권 변화에 민감
    - 1위와 2위 차이가 9위와 10위 차이보다 훨씬 큼
- Hits@k (k = 1, 3, 10)
  - ✓ 정답이 상위 k 안에 들어간 비율
  - ✓ 직관적이지만 k 경계 밖의 차이는 못 봄
    - k=10 일 때 11위나 6,000위나 동일



# CONTENTS

1. Background
2. Approach
  - Embedding-based
  - Rule-based
3. Why LLM?
4. LLM Integration Strategies
5. Research Direction



# Approach

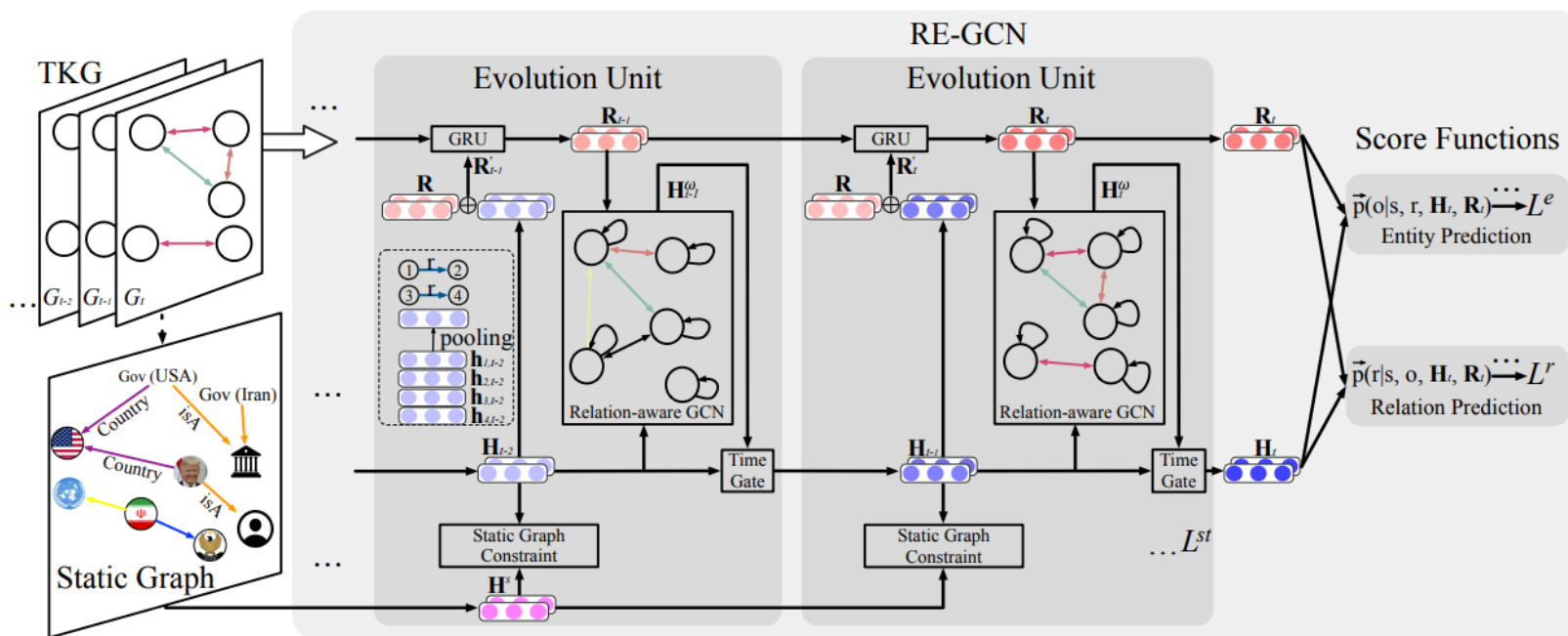
## Plain Embedding-based Model

- Method
  - ✓ History Encoding 구조적 의존성 + 순차 패턴
    - 최근 m개 snapshot을 통해 entity, relation 표현을 학습 및 갱신
  - ✓ Scoring Decoder에 query 표현을 넣어 모든 후보와 비교
    - 학습된 표현을 바탕으로 query에 대한 전체 entity 점수 계산
- Pros
  - ✓ 그래프 신경망을 통해 entity간 상호작용 패턴을 모델링하여 구조적 의존성 반영
  - ✓ 모든 쿼리에 대해 항상 전체 entity 점수를 낼 수 있음
- Cons
  - ✓ 답의 근거를 알 수 없음 (interpretability)
  - ✓ Unseen entity는 표현 자체가 없음
- 대표 논문
  - ✓ [2021][SIGIR][RE-GCN]

# Plain Embedding-based Model

[2021][SIGIR][RE-GCN]

- Key Idea
  - ✓ TKG를 KG snapshot의 sequence로 보고 표현을 시간에 따라 진화
- 핵심 구조 : Evolution Unit
  - ✓ Snapshot 내부 : relation-aware GCN으로 동시 발생 facts들의 구조적 의존성 포착
  - ✓ Snapshot 간 : time gate + GRU로 시간에 따른 표현 변화 파악



# Approach

## Plain Rule-based Model

- Method
  - ✓ Rule mining
  - ✓ Confidence 추정
  - ✓ Rule application
    - Query에 rule을 매칭 -> body를 만족하는 후보 entity에 rule confidence 집계 -> ranking
- Pros
  - ✓ 왜 그 답인지 rule로 설명 가능
  - ✓ Rule이 새 entity에도 적용 가능 (inductive)
- Cons
  - ✓ Rule이 매칭 안 되는 쿼리는 답을 내지 못 함
  - ✓ Relation이 조금만 달라도 rule을 찾지 못함
  - ✓ 통계적으로 판정된 rule이 실제로 말이 되는 rule인지 구분 못 함
- 대표 논문
  - ✓ [2022][AAAI][TLogic]

# Plain Rule-based Model

[2022][AAAI][TLogic]

- Key Idea
  - ✓ History에서 temporal logical rule을 캐내고 그 rule을 query에 적용해 예측
  - ✓ Embedding과 달리 설명 가능한 근거(rule + 실제 경로)를 함께 제공
- Cyclic temporal rule Cyclic : body의 경로가 head와 같은 두 entity를 잇는 다른 경로
  - ✓ Body (조건) -> Head (결론)
    - Ex) (E1, discuss by telephone, E2, T1) -> (E1, consult, E2, T2)
  - ✓ Time constraints : body의 timestamp들이 head보다 앞서야 함
- 핵심 구조 : Rule learning -> Rule application
  - ✓ Rule Learning
    - Temporal random walk로 history에서 시간을 거슬러 경로 sampling
    - 그 walk를 변수로 추상화 -> rule로 변환 만약 walk 결과가 cyclic이 아니면 그 walk는 버려짐
    - 각 rule에 confidence 부여
  - ✓ Rule Application
    - Query relation의 rule들을 confidence 높은 순으로 적용
    - Body를 만족하는 실제 경로를 찾아 후보 entity 추출 각 후보 점수 = rule confidence + 시간의 근접성



# CONTENTS

1. Background
2. Approach
3. Why LLM?
4. LLM Integration Strategies
5. Research Direction

# Why LLM?

## Limitation

- **Symbolic grounding의 부재** 가장 근본적인 한계
  - ✓ Entity를 의미 없는 기호로 다뤄 그 기호의 그래프 내 통계로만 의미를 부여 받음
- Closed-world
  - ✓ 기존 방법들은 그래프에 있는 것이 정보의 전부
  - ✓ 그래프 밖의 상식이나 역사는 존재하지 않음
- Cold-start / Long-tail
  - ✓ Sparse한 entity, relation에서 붕괴
  - ✓ 하지만 sparse한 경우가 forecasting의 가치가 매우 큼
- Inductive 어려움
  - ✓ Unseen entity는 embedding 표현 자체가 없음
- Recurrence bias
  - ✓ 성능 상당 부분이 과거의 fact가 또 나온다는 것에 의존
- No semantic validation Rule-based
  - ✓ 자주 맞은 rule이 말이 되는 규칙인지 구분 못 함

# Why LLM?

## LLM (Large Language Model)

- 대규모 텍스트로 pre-trained된 언어 모델
- In TKG?
  - ✓ Relation 간 선후 전개 패턴을 내재된 지식으로 알 수 있음
    - Ex) "밥먹다" -> "양치하다"는 자연스러운 순서임을 LLM이 알고 있음
  - ✓ Entity, relation을 기호가 아니라 텍스트의 의미로 읽어 정확히 같은 패턴이 아니더라도 유사한 상황으로 일반화 가능
    - History가 없는 unseen entity도 대응 가능
  - ✓ Rule이 자연어로 읽히고 쓰여 LLM과 결합이 자연스러움

# Why LLM?

## Integration Strategies

- LLM-based
  - ✓ 별개의 방법론이 아니라 기존 파이프라인의 특정 지점에 투입되는게 일반적
- Integration Strategies
  - ✓ In-Context Learning
  - ✓ Rule-mediated
  - ✓ Fine-tuning



# CONTENTS

1. Background
2. Approach
3. Why LLM?
4. LLM Integration Strategies
  - In-Context Learning
  - Rule-mediated
  - Fine-tuning
5. Research Direction

## In-Context Learning

LLM 파라미터를 건드리지 않고 history를 프롬프트로 넣어 예측

## **Temporal Knowledge Graph Forecasting Without Knowledge Using In-Context Learning**

**Dong-Ho Lee\*, Kian Ahrabian\*, Woojeong Jin, Fred Morstatter, Jay Pujara**

Department of Computer Science and Information Sciences Institute  
University of Southern California

{dongho.lee, ahrabian, woojeong.jin}@usc.edu, {fredmors, jpujara}@isi.edu

# In-Context Learning

[2023][EMNLP][ICL]

- Key Idea
  - ✓ TKG Forecasting을 fine-tuning 없이 In-Context Learning만으로 접근
  - ✓ History를 text prompt로 만들어 LLM의 다음 토큰 확률로 답을 ranking
- 이후 LLM-based 모델들의 공통 baseline

Which team will win the Super Bowl in 2023?

2000: [Superbowl, Champion, St Louis]  
2001: [Superbowl, Champion, Baltimore]  
2002: [Superbowl, Champion, New England]  
2003: [Superbowl, Champion, Tampa Bay]  
...  
2019: [Superbowl, Champion, New England]  
2020: [Superbowl, Champion, Kansas City]  
2021: [Superbowl, Champion, Tampa Bay]  
2022: [Superbowl, Champion, Los Angeles]  
2023: [Superbowl, Champion,

# In-Context Learning

[2023][EMNLP][ICL]

- History Modeling
  - ✓ Query 관련 history를 history snapshot에서 두 독립적인 축을 기준으로 선택
  - ✓ Entity vs. Pair
    - Entity : query의 subject를 포함하는 모든 과거 fact
    - Pair : subject와 relation을 둘 다 포함하는 과거 fact
  - ✓ Unidirectional vs. Bidirectional
    - Unidirectional : query에서와 같은 위치(subject 또는 object)에 있는 fact만
    - Bidirectional : 위치 상관 없이 fact 포함 단, object에 위치할 경우 inverse quadruple로 방향 통일

|              |             |
|--------------|-------------|
| Entity / Uni | Entity / Bi |
| Pair / Uni   | Pair / Bi   |

# In-Context Learning

[2023][EMNLP][ICL]

- Prompt Construction

- ✓ 추출한 history를 미리 정의된 template으로 각 fact를 나열하고 query는 답을 비워 붙임
- ✓ Index 방식
  - Entity, relation을 숫자 인덱스로 표현
  - Semantic 정보를 제거해 패턴만으로 가능한지 테스트
- ✓ Lexical 방식
  - Entity, relation을 실제 텍스트로 표현

| Category                        | Prompt                                    |
|---------------------------------|-------------------------------------------|
| Lexical<br>$\mathcal{L}(\cdot)$ | 2000: [Superbowl, Champion, 0. St Louis]  |
|                                 | 2001: [Superbowl, Champion, 1. Baltimore] |
|                                 | ...                                       |
|                                 | 2023: [Superbowl, Champion,               |
| Index<br>$\mathcal{I}(\cdot)$   | 2000: [0, 0, 0. 0]                        |
|                                 | 2001: [0, 0, 1. 1]                        |
|                                 | ...                                       |
|                                 | 2023: [0, 0,                              |



# In-Context Learning

[2023][EMNLP][ICL]

- Response Generation
  - ✓ LLM이 생성할 entity 이름이 여러 토큰으로 쪼개져 직접 확률 계산이 어려움
  - ✓ Indirect Logit
    - LLM에 프롬프트를 넣고 다음 numerical label 토큰의 확률을 계산
    - 이름의 확률이 아닌 그 이름이 배정된 숫자의 확률을 뽑도록 함
- Prediction Setting
  - ✓ Single-step
    - Test 기간의 과거 ground truth를 history로 제공
    - 예측 후 현재 timestamp의 실제 답을 다음 history에 추가
  - ✓ Multi-step
    - Test 기간 ground truth가 아닌 자신의 예측 결과를 다음 history에 추가
    - 진행할수록 불확실성 누적

## **AnRe: Analogical Replay for Temporal Knowledge Graph Forecasting**

**Guo Tang<sup>1</sup>, Zheng Chu<sup>1</sup>, Wenxiang Zheng<sup>1</sup>, Junjia Xiang<sup>1</sup>, Yizhuo Li<sup>1</sup>,  
Weihao Zhang<sup>1</sup>, Ming Liu<sup>1,2\*</sup>, Bing Qin<sup>1,2</sup>**

<sup>1</sup>Harbin Institute of Technology, Harbin, China

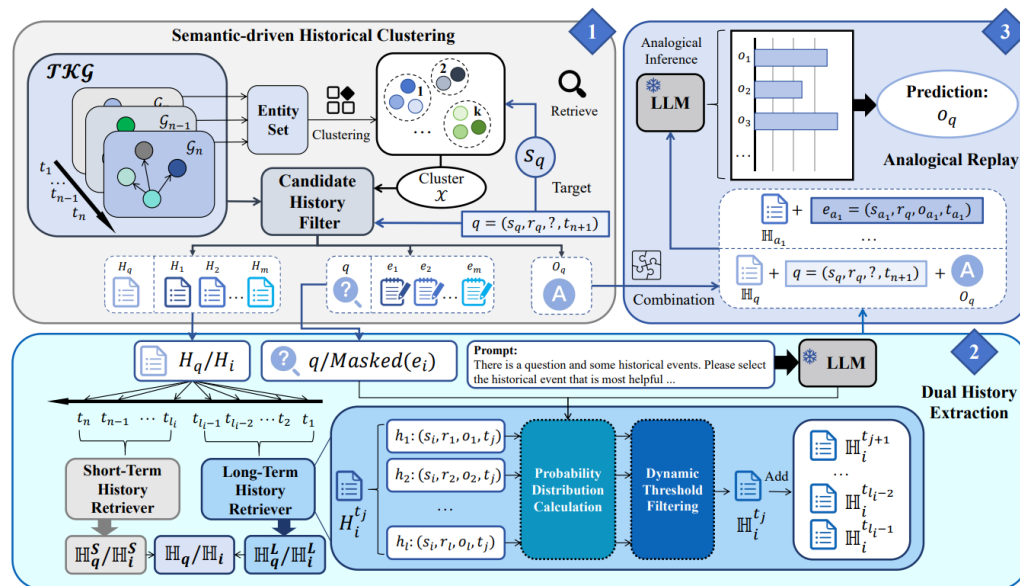
<sup>2</sup>Peng Cheng Laboratory, Shenzhen, China

{gtang,zchu,wxzheng,mliu,qinb}@ir.hit.edu.cn

# In-Context Learning

[2025][ACL][AnRe]

- 기존 ICL의 Limitation
  - ✓ Zero-shot 추론에만 의존
  - ✓ Fact를 자연어로 풀어 쓰지 않아서 의미 이해가 약함
- Key Idea
  - ✓ Training-free로 In-Context Learning을 정교화
  - ✓ Query의 history가 부족해도 유사한 사건의 과거 전개를 빌려와 예측



# In-Context Learning

[2025][ACL][AnRe]

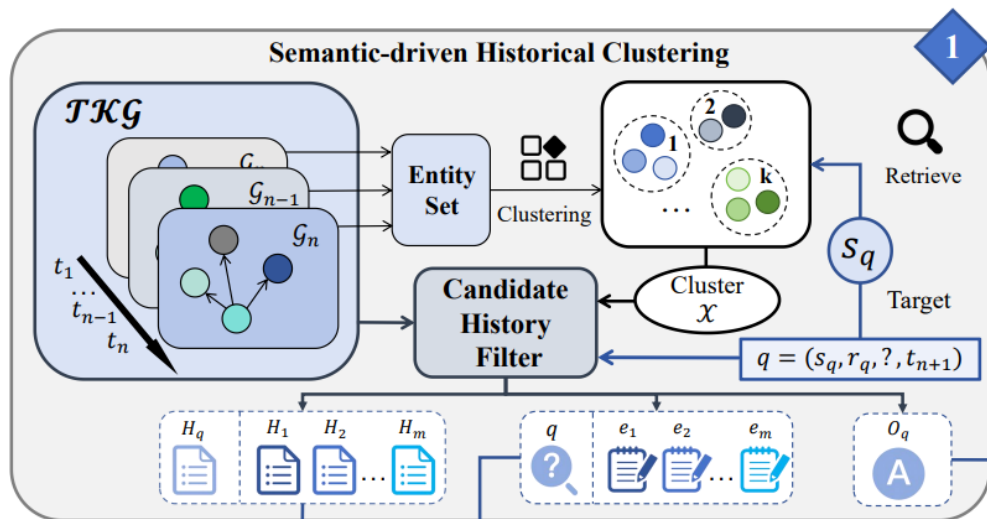
- Semantic-driven Historical Clustering

- ✓ Entity Semantic Clustering

- 전체 entity를 BERT로 embedding
- K-means clustering 후 query entity  $s_q$ 가 속한 클러스터  $\mathcal{X}$ 를 유사 엔티티 집합으로 선택

- ✓ Candidate History Filter

- 클러스터 내 각 유사 entity에 대해 query relation을 갖는 history  $H_i$  추출
- History가 300개 이상인  $H_i$ 와  $e_i$ 만 남기고 유사도 순으로 정렬 ( $e_i$ : 가장 유사한 event)
- $s_q$ 에 대해서도 관련 history  $H_q$ 와 candidate answer 집합  $O_q$  구성



# In-Context Learning

[2025][ACL][AnRe]

- Dual History Extraction

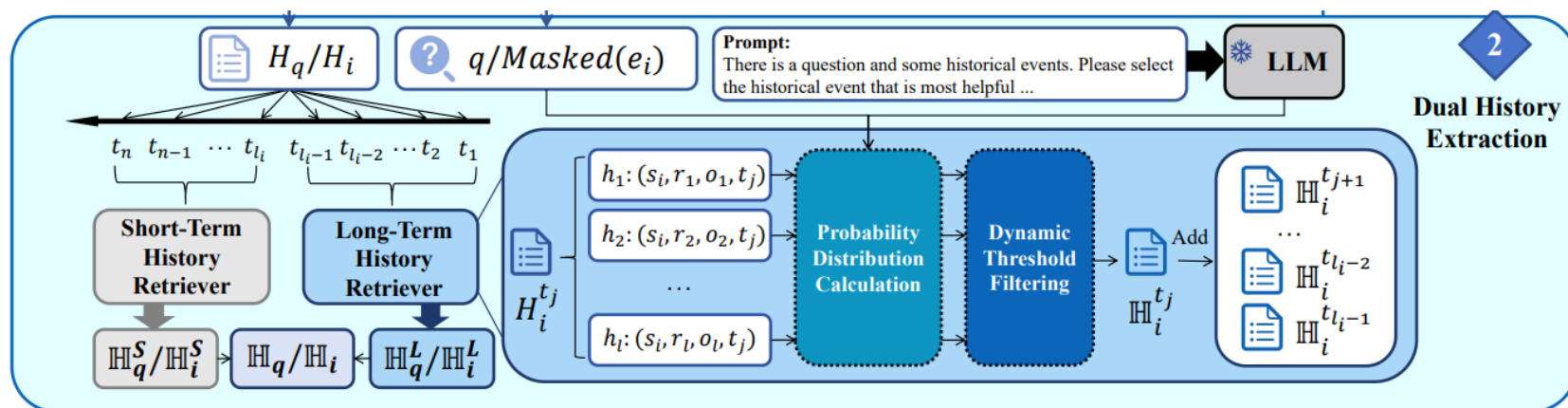
- ✓ Short-term history retriever

- Query와 유사 사건  $e_i$ 의 관련 history  $H_i$ 를 timestamp 순으로 정렬
- 가장 최근  $l$ 개를 단기 체인  $\mathbb{H}_S^i$ 로 만들어 직접적인 인과와 급변을 포착

- ✓ Long-term history retriever

- Probability Distribution Calculation : LLM에게 각 과거 사건이 query 추론에 얼마나 유용한지 판단 시켜 사건별 유효성 확률 산출
- Dynamic Threshold Filtering : 시간 차이 기반 동적 임계값으로 유효 history만 남김

- ✓ 하나의 history chain으로 통합

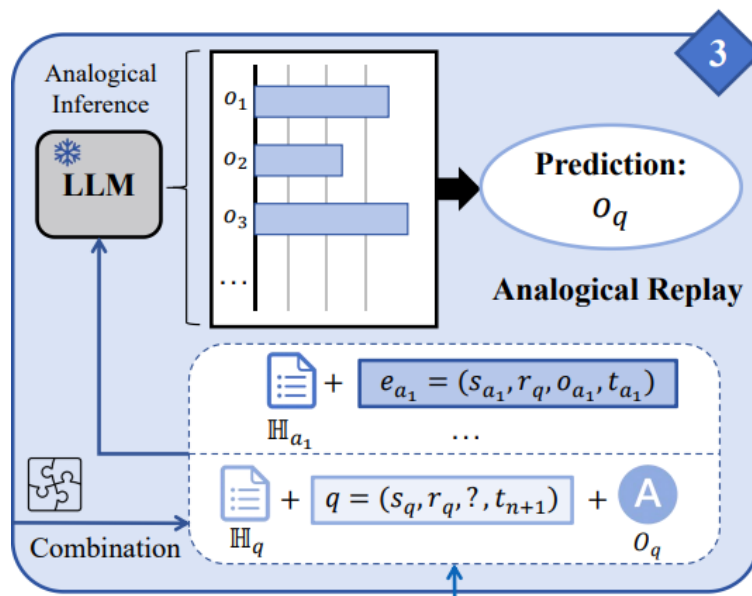


# In-Context Learning

[2025][ACL][AnRe]

- Analogical Replay

- ✓ 유사 event와 그 history를 LLM에게 입력하여 history 맥락에서 어떻게 event의 답이 도출되는지 reasoning 과정 생성
  - 유추 예시 구성 : history + 유사 event + LLM 분석
- ✓ Query와 가장 유사한 상위 a개 event의 유추 예시 + query의 history + 후보 entity를 프롬프트로 넣어 LLM이 최종 예측
- ✓ 후보를 numerical token으로 매핑해 확률분포로 ranking Indirect logit



## Rule-mediated

LLM으로 logical rule을 생성 / 정제한 뒤 그 rule로 추론하는 방식

---

## Large Language Models-guided Dynamic Adaptation for Temporal Knowledge Graph Reasoning

---

**Jiapu Wang<sup>1</sup>, Kai Sun<sup>1\*</sup>, Linhao Luo<sup>2</sup>, Wei Wei<sup>3</sup>, Yongli Hu<sup>1†</sup>, Alan Wee-Chung Liew<sup>4</sup>,  
Shirui Pan<sup>4†</sup>, Baocai Yin<sup>1</sup>**

<sup>1</sup>Beijing University of Technology, China, <sup>2</sup>Monash University, Australia

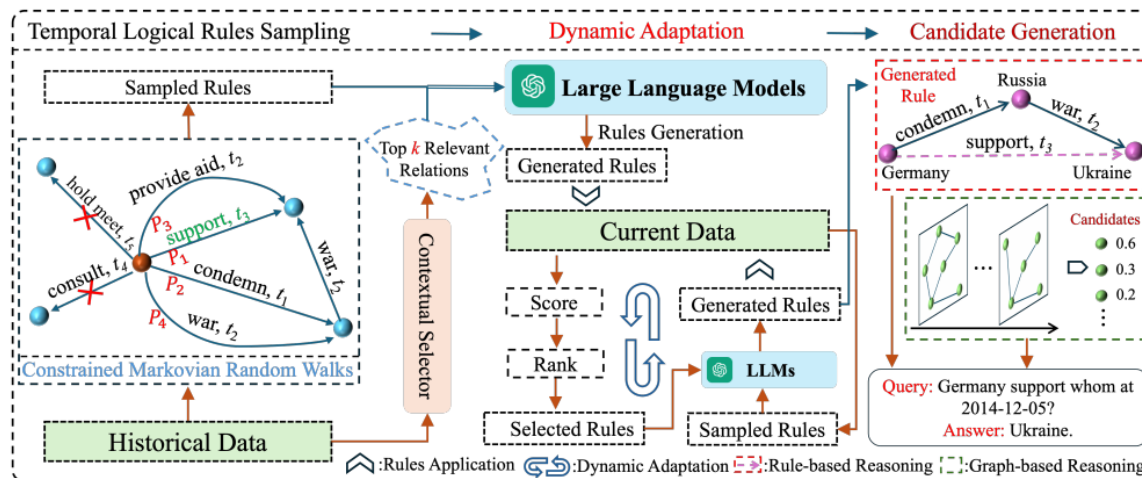
<sup>3</sup>University of Hong Kong, China, <sup>4</sup>Griffith University, Australia

{jpwang, sunkai}@emails.bjut.edu.cn, linhao.luo@monash.edu, weiwei1206cs@gmail.com  
{huyongli, ybc}@bjut.edu.cn, {a.liew, s.pan}@griffith.edu.au

# Rule-mediated

[2024][NIPS][LLM-DA]

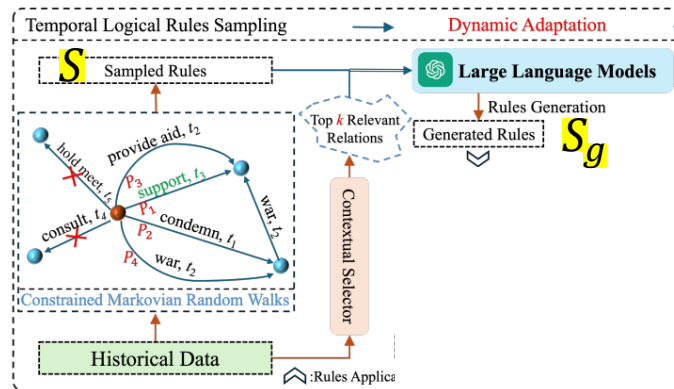
- Limitation
  - ✓ Plain rule-based model은 새로운 지식에 맞게 update 시키는 것 자체가 난제
  - ✓ LLM-based model은 LLM 자체가 blackbox여서 어떤 temporal pattern이 추론에 기여했는지 불투명
- Key Idea
  - ✓ LLM으로 명시적인 temporal logical rule을 생성해 추론 과정을 interpretability하게 만듦
  - ✓ 생성한 rule을 최신 데이터로 iterative하게 갱신
  - ✓ 각 rule head와 의미적으로 가까운 top-k relation만 골라 context로 LLM에 넣어 rule 생성



# Rule-mediated

[2024][NIPS][LLM-DA]

- Temporal Logical Rules Sampling
  - ✓ Constrained Markovian random walk
    - 두 가지 제약
      - Temporal order : 나열되는 body 전체가 head보다 먼저 일어난 사건이도록 강제
      - Temporal interval : 전이 확률에 exponential decay를 적용해 최근 edge에 높은 가중치
    - 결과로 규칙 집합  $S$ 를 얻음
  - ✓ Rule Generation
    - Contextual Relation Selector가 Sentence-BERT로 rule-head와 후보 relation을 임베딩하고 cosine similarity로 top-k relation 선택
    - 추출된  $S$ 와 top-k relation을 LLM에 넣어 일반 규칙  $S_g$ 를 생성

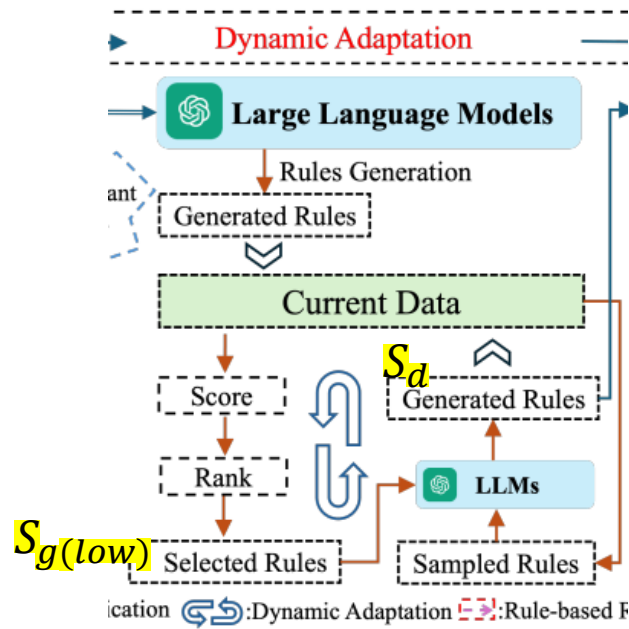


# Rule-mediated

[2024][NIPS][LLM-DA]

- Dynamic Adaption

- ✓ 일반 규칙들 중 Current Data 위에서 confidence가 threshold보다 낮은 규칙  $S_{g(low)}$  식별
- ✓ 동시에 Current Data에서도 constrained random walk로 규칙을 뽑아 이를 기준으로 삼고 LLM이 저품질 규칙  $S_{g(low)}$ 를 갱신해 규칙 집합  $S_d$ 를 생성
- ✓ 이 과정을 반복



# Rule-mediated

[2024][NIPS][LLM-DA]

- Candidate Reasoning

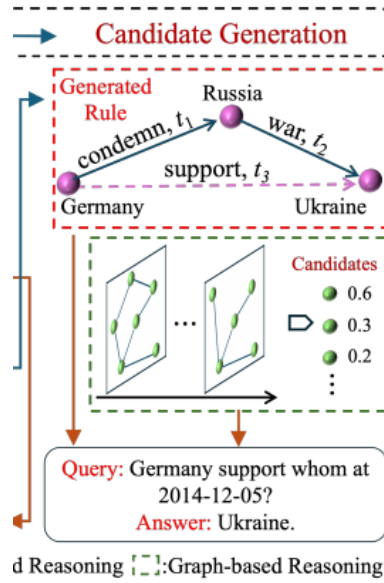
- ✓ Rule-based scoring

- High confidence rule들로 추론 경로를 찾아 후보 entity를 도출하고 confidence + time decay로 scoring

- ✓ Graph-based scoring

- Query와 entity들의 임베딩 내적으로 scoring

- ✓ 최종 점수는  $\text{Score}_f = \alpha \cdot \text{Rule score}_{(\rho, e'_o)} + (1 - \alpha) \cdot \text{Graph score}_{(\text{graph}, e'_o)}$



## **Critic Rule Induction: Improving Temporal Knowledge Graph Forecasting with Generator-Critic Language Models**

**Yingsong Ning, Fu Zhang<sup>\*</sup>, Jingwei Cheng, Jiashun Peng, Xiaoke Wang**

School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

{2401887, 2472039, 2401911}@stu.neu.edu.cn

{zhangfu, chengjingwei}@neu.edu.cn

# Rule-mediated

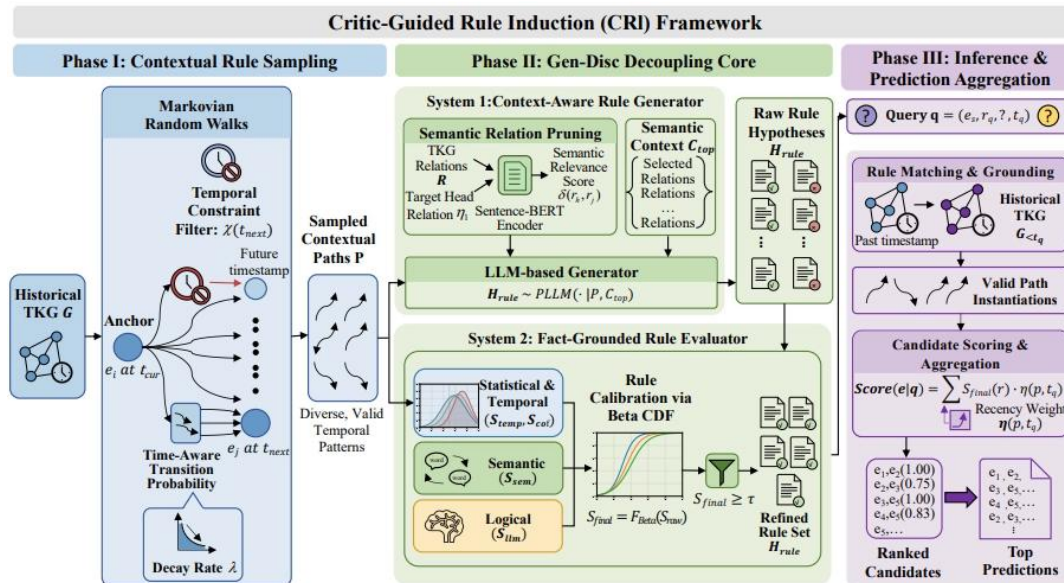
[2026][ACL-F][CRI]

- Limitation

- ✓ Rule confidence를 단일 기준 하나로만 평가
  - sparse relation은 rule이 만들어지지 않거나 만들어져도 커버하는 query가 극소수
  - 기존 조건부 확률로 되어있는 confidence는 인과 없이 우연히 자주 같이 나타나는 두 사건에서도 높게 나옴

- Key Idea

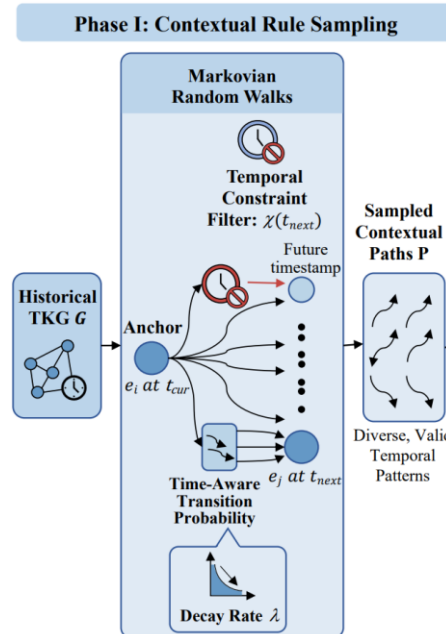
- ✓ Temporal rule을 검증 대상인 hypothesis로 취급하고 Generation-Discrimination을 분리



# Rule-mediated

[2026][ACL-F][CRI]

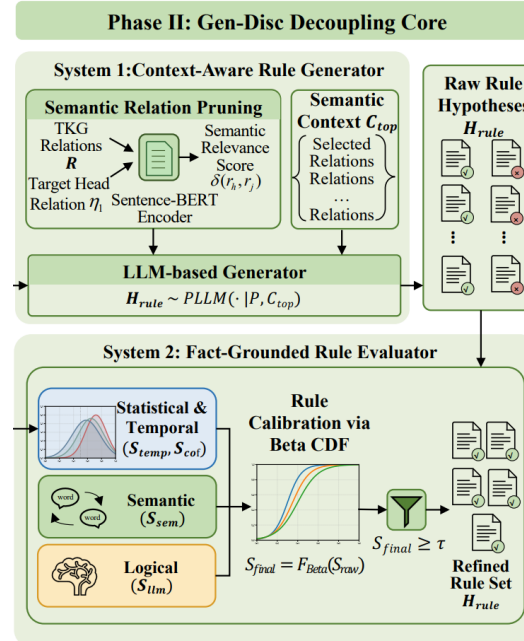
- Contextual Rule Sampling Anchor node에서 시작
  - ✓ Temporal Constraint Filter
    - 다음 이동할 event의 시점이 현재 시점보다 이전일 때만 통과
  - ✓ Time-Aware Transition Probability
    - Filter를 만족하는 이웃들에 대해서 시간적으로 가까운 fact들을 선호하도록 확률 정의
  - ✓ 이를 바탕으로 path 집합  $\mathcal{P}$ 를 얻음



# Rule-mediated

[2026][ACL-F][CRI]

- Gen-Disc Decoupling Core
  - ✓ Context-Aware Rule Generator
    - Rule을 만들 target relation과 의미적으로 관련 있는 relation set  $\mathcal{C}_{top}$  구성
    - 앞에서 얻은 path 집합  $\mathcal{P}$ 와  $\mathcal{C}_{top}$ 를 LLM에 입력하여 rule 생성
  - ✓ Fact-grounded Rule evaluator
    - Discriminator가 rule을 네 가지 차원으로 평가



# Rule-mediated

[2026][ACL-F][CRI]

- Gen-Disc Decoupling Core
  - ✓ Temporal Constraint Score  $S_{temp}(r)$ 
    - Strict temporal constraint를 만족하는 path의 비율을 계산
  - ✓ Historical Confidence Score  $S_{conf}(r)$ 
    - 기존 rule confidence 계산 방식과 동일 (conditional probability)
  - ✓ Semantic Consistency Score  $S_{sem}(r)$ 
    - high-quality 템플릿과 low-quality 템플릿을 이용해 contrastive similarity 방법으로 점수 부여
  - ✓ LLM-based Logical Coherence Score  $S_{llm}(r)$ 
    - 임베딩이 놓칠 수 있는 논리적 불일치를 잡기 위해 LLM으로 평가
  - ✓ 위 결과들을  $S_{raw}$ 로 aggregate한 뒤 high quality rule과 low quality rule을 더 극명하게 구분하기 위해  $S_{final}(r)$ 정의 => refined rule set  $\mathcal{H}'_{rule}$

$$\begin{aligned} S_{final}(r) &= F_{Beta}(S_{raw}(r); \alpha, \beta) \\ &= \int_0^{S_{raw}(r)} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} dx \end{aligned}$$

# Rule-mediated

[2026][ACL-F][CRI]

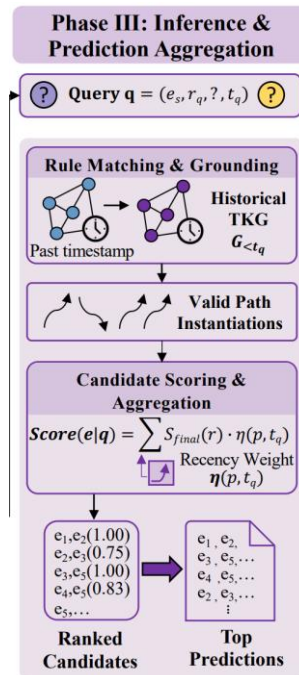
- Inference & Prediction Aggregation

- ✓ Rule matching

- Query의 relation에 매칭되는 rule을  $\mathcal{H}'_{rule}$ 에서 retrieve하고 이를 과거 graph에 ground 시킴

- ✓ Candidate Scoring and Aggregation

- 유효한 path들이 가리키는 candidate entity를 여러 rule과 여러 path의 기여를 합산해 scoring



## Fine-tuning

TKG 데이터로 LLM을 추가 학습시켜 도메인에 적응시키는 방식

## **GenTKG: Generative Forecasting on Temporal Knowledge Graph with Large Language Models**

**Ruotong Liao<sup>1,2</sup>, Xu Jia<sup>3</sup>, Yangzhe Li<sup>3</sup>, Yunpu Ma<sup>1,2,4</sup>, Volker Tresp<sup>1,2</sup>**

<sup>1</sup>LMU Munich    <sup>2</sup>Munich Center for Machine Learning (MCML)

<sup>3</sup>Technical University of Munich    <sup>4</sup>Siemens AG

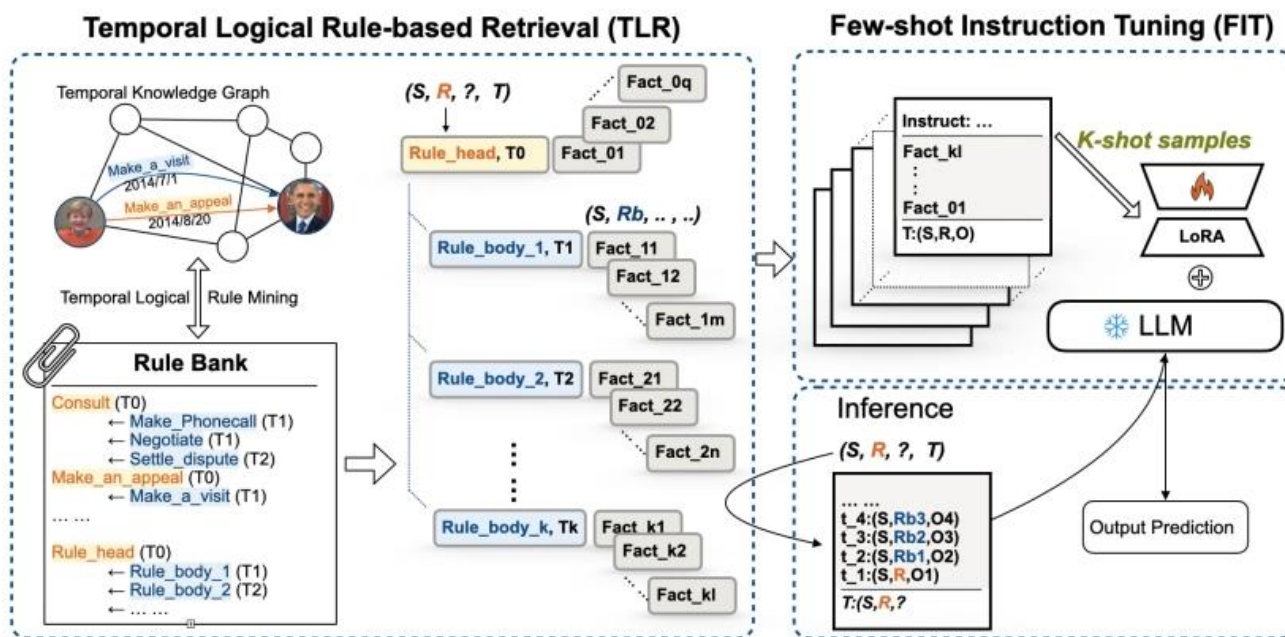
ruotong.liao@outlook.com,    cognitive.yunpu@gmail.com

volker.tresp@lmu.de

# Fine-tuning

## [2024][NAACL-F][GenTKG]

- Limitation
  - ✓ 기존 모델들은 데이터셋의 패턴을 모델에 학습 시킴 *data-centric learning*
- Key Idea
  - ✓ Rule을 이용해 query와 관련된 history를 자연어로 표현해 LLM에게 제공
  - ✓ LLM에게 데이터의 패턴을 학습시키지 않고 task 자체에 align *task-centric alignment*



# Fine-tuning

## [2024][NAACL-F][GenTKG]

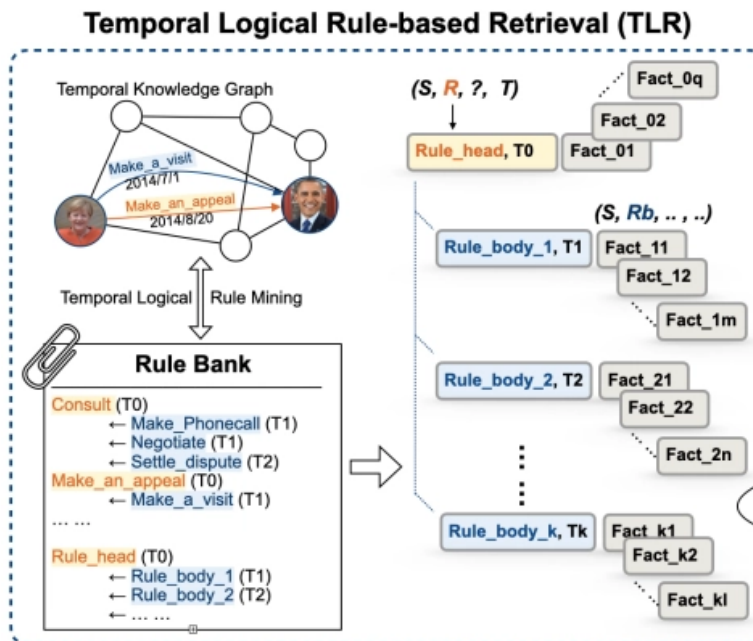
- Temporal Logic Rule-based Retrieval

- ✓ Rule mining

- TLogic을 이용해 temporal logical rule mining해 rule bank 구성

- ✓ Retrieval

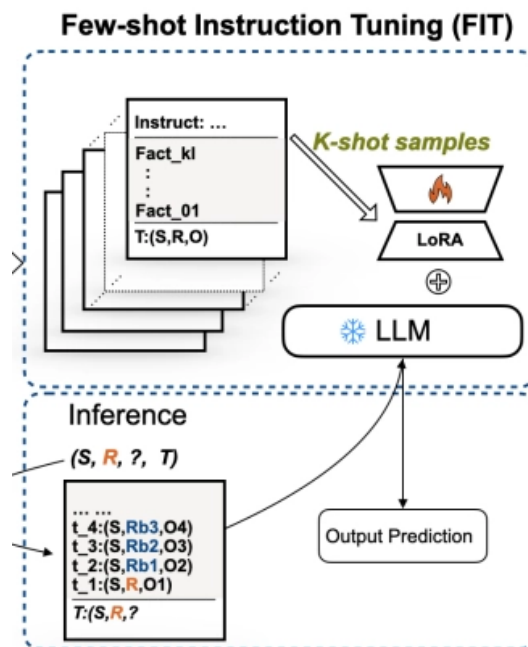
- Query relation을 rule head로 간주하고 관련된 history fact를 가져옴
- Rule을 적용해 rule body들을 confidence 내림차순으로 top-k 선택하고 body에 해당하는 과거 event를 추가



# Fine-tuning

## [2024][NAACL-F][GenTKG]

- Few-shot parameter-efficient Instruction Tuning
  - ✓ Instruction prompt
    - Task instruction + TLR facts + task output 으로 구성
    - Entity name을 직접 사용해 예측
  - ✓ PEFT
    - LoRA로 파라미터 효율적 튜닝



## **RECIPE-TKG: From Sparse History to Structured Reasoning for LLM-based Temporal Knowledge Graph Completion**

**Ömer Faruk Akgül<sup>1\*</sup> Feiyu Zhu<sup>1\*</sup> Yuxin Yang<sup>1</sup> Rajgopal Kannan<sup>2</sup> Viktor Prasanna<sup>1</sup>**

<sup>1</sup>University of Southern California

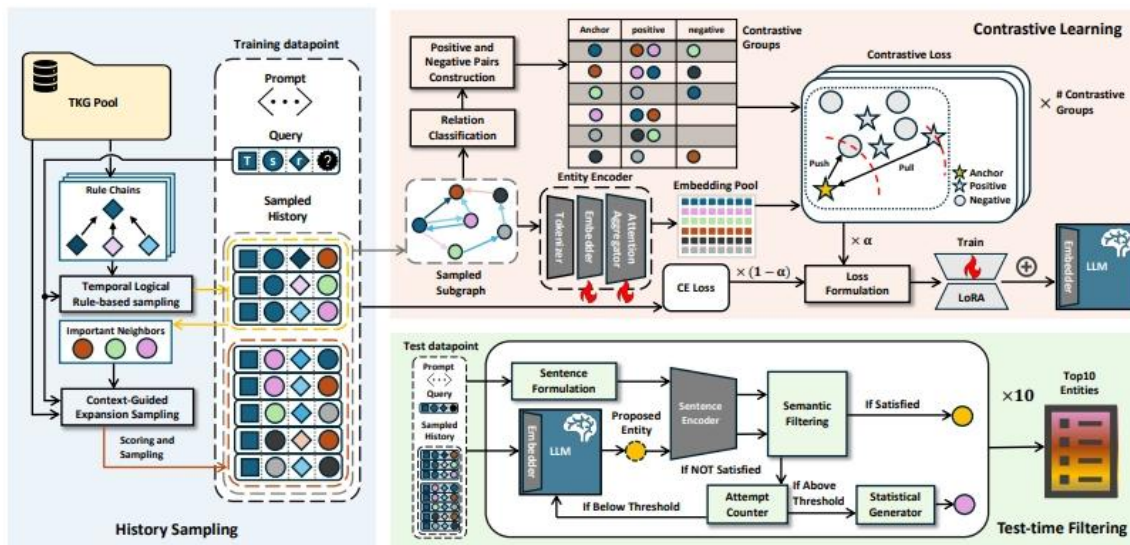
<sup>2</sup>DEVCOM Army Research Office

{akgul, feiyuzhu, yyang393, prasanna}@usc.edu, rajgopal.kannan.civ@army.mil

# Fine-tuning

## [2026][EACL][RECIPE-TKG]

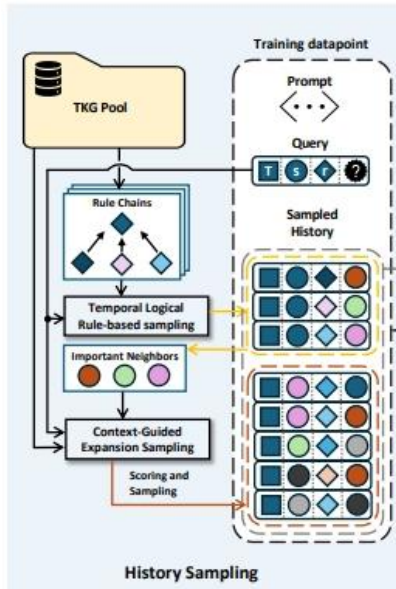
- Limitation
  - ✓ 기존 retrieval은 1-hop history에 의존
  - ✓ 표준 fine-tuning이 일반화 대신 memorization을 강화함
  - ✓ Uncalibrated 생성이 implausible entity를 생성할 수 있음
- Key Idea
  - ✓ 위 세 가지 limitation을 세 모듈이 보강
    - Multihop sampling / contrastive fine-tuning / semantic filtering



# Fine-tuning

## [2026][EACL][RECIPE-TKG]

- History Sampling : Rule-based multi-hop retrieval (RBMH)
  - ✓ Temporal Logical Rule-based Sampling
    - GenTKG의 TLR로 1-hop fact mining
  - ✓ Context-guided Multi-hop Expansion
    - 1-hop fact를 채우고 LLM의 context가 남은 경우 multi-hop 후보들의 weight를 계산
    - weight에 비례한 확률로 최종 history 구성



# Fine-tuning

## [2026][EACL][RECIPE-TKG]

- Composite Weight

$$w = w_n \cdot w_f \cdot (w_t + w_c + w_{cp}),$$

- ✓ Neighbor weight (hard filter) Hop 거리가 멀수록 감점

$$w_n = \exp(-\gamma_1 \cdot (\text{hop}_s + \text{hop}_o - 1))$$

- $\text{hop}_s, \text{hop}_o$  : Query subject로부터 후보 fact의 subject / object까지의 shortest hop distance가 클수록 지수적으로 감쇠하는 값
- 둘 중 하나라도  $\infty$  (도달 불가) -> neighbor weight = 0 으로 후보에서 완전 제거

- ✓ Frequency weight 자주 등장하는 triple 감점

$$w_f = \frac{1}{\gamma_2 \cdot \log(n_{spo}) + 1}$$

- $n_{spo}$  : timestamp를 제외한  $(s, p, o)$  triple의 등장 횟수
- 더 희귀한 triple이 더 높은 weight

# Fine-tuning

## [2026][EACL][RECIPE-TKG]

- Composite Weight

$$w = w_n \cdot w_f \cdot (w_t + w_c + w_{cp}),$$

- ✓ Time weight 최근 사건 우선

$$w_t = \exp \left( -\gamma_3 \cdot \frac{T - t}{\delta} \right)$$

- $\delta$  : time granularity (timestamp 기준)
- Query에 시간적으로 가까운 후보에 더 큰 가중치를 줌

- ✓ Frequency weight Query subject / relation과 동시 발생 횟수 반영

$$w_c = \frac{\log(1 + \gamma_4 \cdot n_{so})}{1 + \log(1 + \gamma_4 \cdot n_{so})}$$

- $n_{so}$  :  $T$  이전에 subject - object 쌍이 함께 나타난 횟수
- 함께 자주 나타나는 entity 쌍을 우대
- Bounded function 이라서 특정 entity가 과도하게 지배하지 않음

# Fine-tuning

## [2026][EACL][RECIPE-TKG]

- Composite Weight

$$w = w_n \cdot w_f \cdot (w_t + w_c + w_{cp}),$$

- ✓ Contextual priority weight TLR로 찾은 entity와 연결된 경우 가중

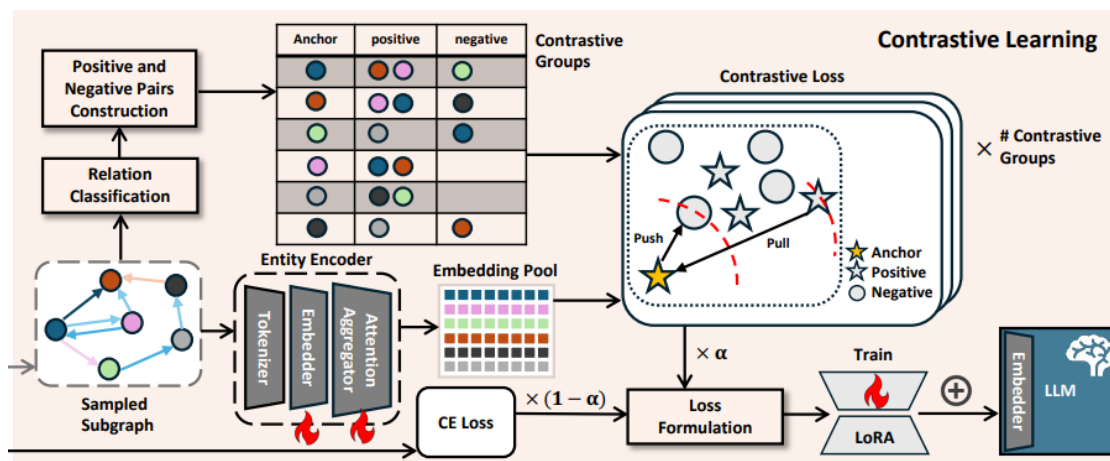
$$w_{cp} = \begin{cases} 1, & \text{if } s \in \mathcal{E}_{\text{TLR}} \text{ or } o \in \mathcal{E}_{\text{TLR}}, \\ 0, & \text{otherwise,} \end{cases}$$

- $\mathcal{E}_{\text{TLR}}$  : temporal logical rule-based sampling에서 식별된 모든 1-hop neighbor 집합
- Multi-hop expansion 0 | 1-hop neighbor와 연결되면 가중치 부여
- 0 | weight는 composite weight에서 덧셈으로 연결되어 0이더라도 후보가 죽지 않음

# Fine-tuning

## [2026][EACL][RECIPE-TKG]

- Contrastive Learning
  - ✓ Relational-Guided Contrastive Pair Construction
    - Relation을 LLM을 사용하여 positive, negative, neutral로 분류
    - RBMH가 뽑은 history에서 각 unique entity의 1-hop neighbor를 훑으며 type 배정
      - Positive relation으로 연결 -> positive sample
      - Negative relation으로 연결 -> negative sample
      - 둘 다 연결된 경우 / neutral relation -> 무시



# Fine-tuning

## [2026][EACL][RECIPE-TKG]

- Contrastive Learning
  - ✓ Training Objective

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{\text{contrastive}} + (1 - \alpha) \cdot \mathcal{L}_{\text{ce}}(o, o_p)$$

- $\mathcal{L}_{\text{ce}}$  : 예측된 token  $o_p$ 와 ground truth  $o$  사이 cross-entropy
- $\alpha$  : balancing hyperparameter
- $\mathcal{L}_{\text{contrastive}}$  : contrastive loss

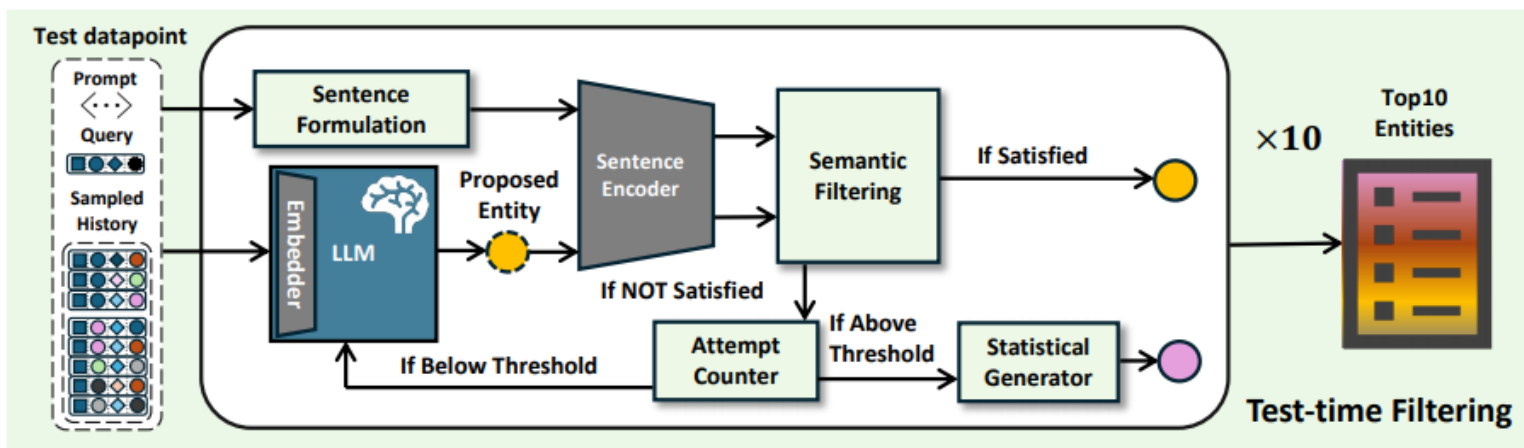
$$\mathcal{L}_{\text{contrastive}} = \frac{1}{N_c} \sum_{i=1}^{N_c} \max \left( 0, \right. \\ \left. \|a_i - pos_i\|^2 - \|a_i - neg_i\|^2 + m \right)$$

- $a_i$  : anchor entity의 embedding
- $pos_i$  : embedding space에서 anchor와 가장 먼 positive
- $neg_i$  : embedding space에서 anchor와 가장 가까운 negative
- Positive와 가까워지고 negative와 멀어지도록 하는 학습 방법

# Fine-tuning

## [2026][EACL][RECIPE-TKG]

- Similarity-Based Test-Time Filtering
  - ✓ 생성 단계에서 말이 안 되는 답을 filtering 하며 반복 정제
  - ✓ LLM이 정답 후보를 생성하면 semantic filter가 예측 결과와 입력 context를 임베딩 유사도로 맥락 적합성 검증
  - ✓ 부적합하면 statistical generator로 재생성
    - 확신 있는 예측이 나올 때까지 iterative하게 refine





# CONTENTS

1. Background
2. Approach
3. Why LLM?
4. LLM Integration Strategies
5. Research Direction

# Research Direction

## Observation (in progress)

- Rule confidence는 대부분 frequency 기반
  - ✓ 후보 entity를 선택하는 과정에서 과거 fact를 기반으로 계산했던 static confidence를 그대로 사용
    - 예측 시점에서도 신뢰할만한 rule인지 판단하지 않음
  - ✓ 빈도가 낮지만 타당한 패턴이 저평가 됨
    - Entity/relation이 rare한 경우가 오히려 forecasting이 어렵고 의미 있는 지점
- Sparse / cold start entity 문제
  - ✓ Rule은 body grounding에 query subject entity의 history가 필요
    - 이 subject entity의 history가 sparse하면 grounding이 부실해 예측이 어려움
    - 반대로 정답 object entity가 history에서 sparse하면 흔한 entity가 상위 후보가 될 가능성이 높음

# Research Direction

## Direction (in progress)

- Rule confidence의 frequency bias -> 의미적 타당성 보완
  - ✓ Frequency가 놓칠 수 있는 rule 내 relation의 타당성을 보완할 수 있는 scoring 방법 모색
- Sparse entity -> 유사 entity로 grounding 보강
  - ✓ AnRe는 semantic 유사도로 entity를 clustering
    - 단순히 query subject entity를 의미적으로 유사한 entity의 rule을 적용하여 prediction하는 것은 어려울 수 있음
    - Rule을 고려한 유사도 반영이 필요

Rule + embedding + LLM을 결합해 의미적으로 타당하고  
sparse에도 강한 rule을 활용하는 방향



**Thank you for Listening**