[CSEG516/AIE5101]

https://github.com/bizli0618/LUPI_SKD

Building a Cultural Context Aware AI

2025. 12. 07 | 20201082 박현빈 | 120240574 이정빈

Building a Cultural Context Aware AI

1. 개요
   1) Why Cultural Context Matters
   2) Preliminaries: KD
   3) Goal of this Project
2. 환경 및 데이터셋
3. 알고리즘 소개
4. 실험 설계
5. 결과 분석
6. 토의 및 결론

개요 Why Cultural Context Matters

The Challenges

- LRMs(Large Reasoning Models) often fails to generate correct reasoning path when cultural knowledge is required.
- This happens more frequently in smaller LRMs (1.7B, 4B size).

➔ Input Query: 제사상을 차릴 때 홍동백서는 무엇을 의미하나요?

Qwen3-32B	**Qwen3-4B**
홍 could be 빨강, meaning red. 동 is 동서의 동, east. 백 is white, and 서 is west. So, 홍 is red, 동 is east, 백 is white, 서 is west. That makes sense with the four cardinal directions. … "홍동백서(紅東白西)"는 조선시대 예절서인『규례』와 관련된 제사의식에서 사용되는 표현으로, 제사상(祭饌)의 배치 원칙을 설명합니다.	I know that in Korean, "백서" (baekseo) is a term related to a type of paper used in rituals, like a "baekseo" which is a type of white paper used for offerings. … "홍동백서" (홍동백서)는 전통 한국의 제사상(祭壇)을 세우는 과정에서 사용되는 **백서(白書)** 의 한 종류로,

개요 Preliminaries: KD

Knowledge Distillation(KD)

- KD trains a student model M_s to match a teacher model M_t by minimizing a token-level divergence between their output distributions.

Off-Policy KD

- Student is trained only on teacher-generated answers or offline labels.
- This is called as Supervised Fine-Tuning(SFT) in some perspective.
- However, this needs lots(a lot!) of datasets and often destroys the student’s own reasoning ability

On-Policy KD

- Student rolls out its own samples, and the teacher is used to match or correct token-level behavior.
- Speculative KD is one of this type.
- Teacher and student usually see the same input Q only.

개요 Goal of This Project

Can we inject new cultural knowledge while preserving the student’s reasoning path?

- On-Policy KD: 한국 문화 관련지식을 smaller LRM에 주입하는 training framework 제안
- With small amount of dataset, we can train the model but not destroying the original reasoning abilities.

Qwen3-4B (AS-IS)	Qwen3-4B (TO-BE)
I know that in Korean, "백서" (baekseo) is a term related to a type of paper used in rituals, like a "baekseo" which is a type of white paper used for offerings. … "홍동백서" (홍동백서)는 전통 한국의 제사상(祭壇)을 세우는 과정에서 사용되는 **백서(白書)**의 한 종류로,	홍 could be 빨강, meaning red. 동 is 동서의 동, east. 백 is white, and 서 is west. So, 홍 is red, 동 is east, 백 is white, 서 is west. That makes sense with the four cardinal directions. … "홍동백서(紅東白西)"는 조선시대 예절서인 『규례』와 관련된 제사의식에서 사용되는 표현으로, 제사상(祭饌)의 배치 원칙을 설명합니다.

Building a Cultural Context Aware AI

1. 개요
2. 환경 및 데이터셋
    1) Cultural Dataset Construction
    2) State/Action/Reward
3. 알고리즘 소개
4. 실험 설계
5. 결과 분석
6. 토의 및 결론

환경 및 데이터셋 Cultural Dataset Construction

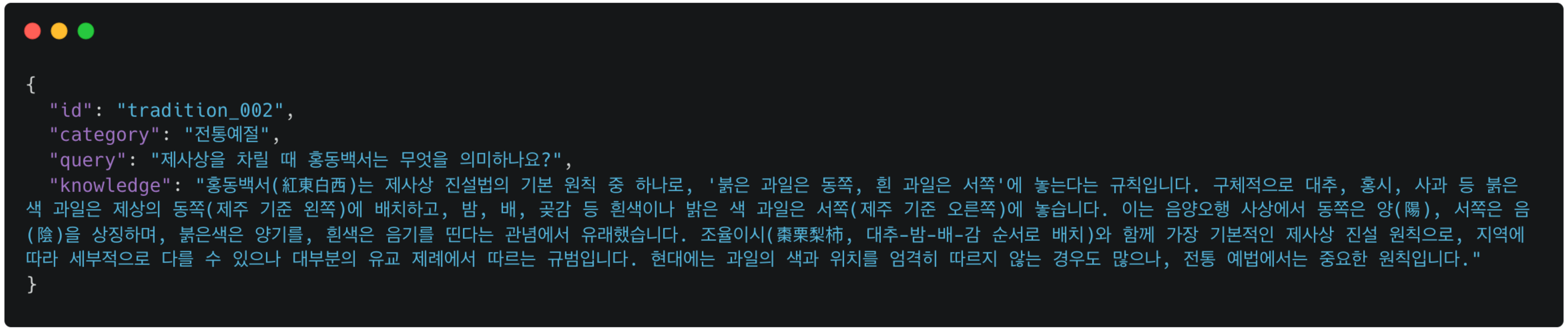
Training Data

- 200 (Query, Knowledge) pairs
  - 5 categories, 40 samples
- Sources are:
  - Encyclopedia of Korean Culture
  - Wikipedia

- Korean, English(machine translated)

Category (EN)	Count
Tradition	40
History	40
Food	40
Geography	40
Modern Culture	40
Total	**200**

- Knowledge field contains a short paragraph describing a specific Korean cultural concept.

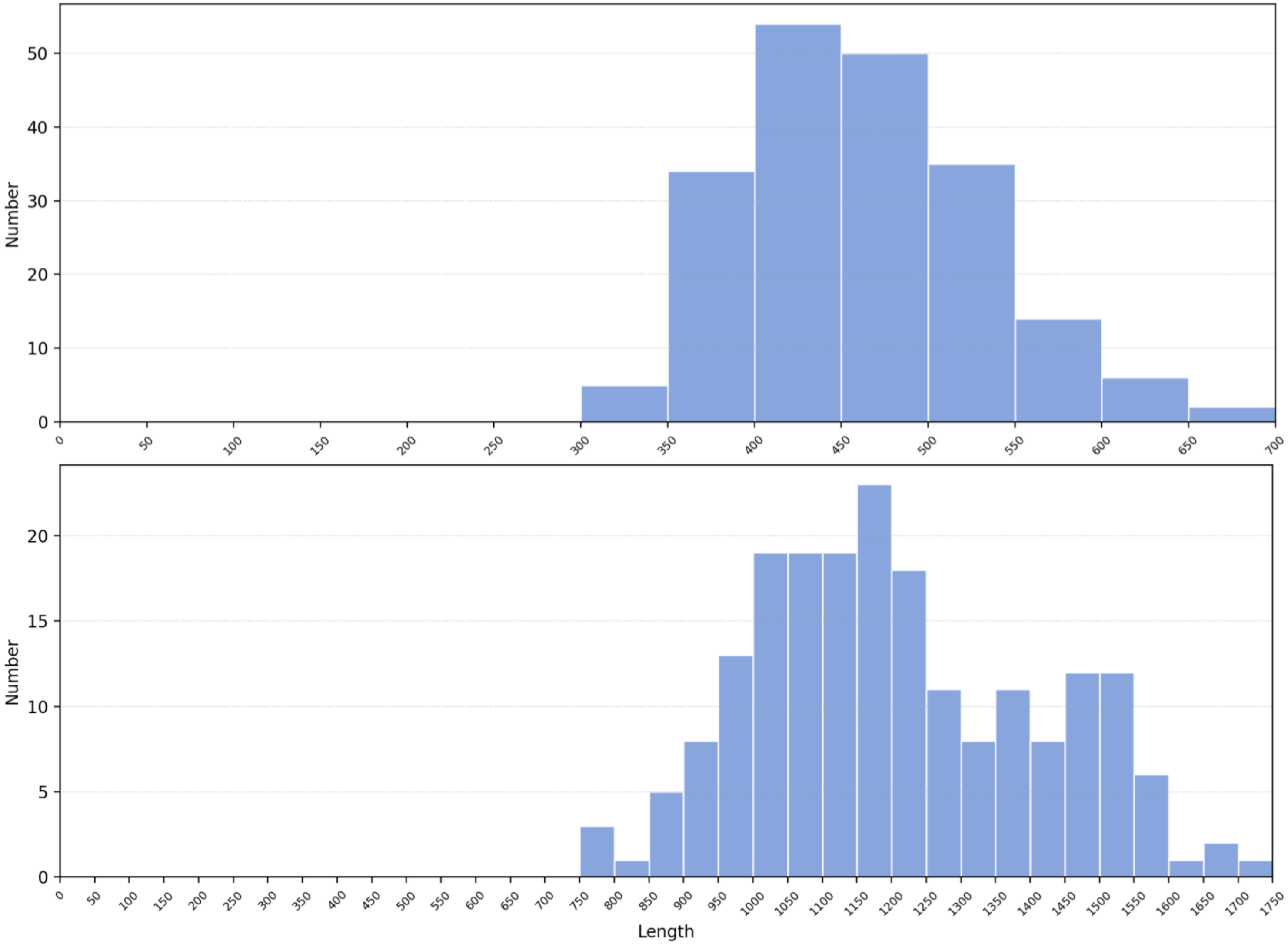


```
{
  "id": "tradition_002",
  "category": "전통예절",
  "query": "제사상을 차릴 때 홍동백서는 무엇을 의미하나요?",
  "knowledge": "홍동백서(紅東白西)는 제사상 진설법의 기본 원칙 중 하나로, '붉은 과일은 동쪽, 흰 과일은 서쪽'에 놓는다는 규칙입니다. 구체적으로 대추, 홍시, 사과 등 붉은
색 과일은 제상의 동쪽(제주 기준 왼쪽)에 배치하고, 밤, 배, 곶감 등 흰색이나 밝은 색 과일은 서쪽(제주 기준 오른쪽)에 놓습니다. 이는 음양오행 사상에서 동쪽은 양(陽), 서쪽은 음
(陰)을 상징하며, 붉은색은 양기를, 흰색은 음기를 띤다는 관념에서 유래했습니다. 조율이시(棗栗梨柿, 대추-밤-배-감 순서로 배치)와 함께 가장 기본적인 제사상 진설 원칙으로, 지역에
따라 세부적으로 다를 수 있으나 대부분의 유교 제례에서 따르는 규범입니다. 현대에는 과일의 색과 위치를 엄격히 따르지 않는 경우도 많으나, 전통 예법에서는 중요한 원칙입니다."
}
```

환경 및 데이터셋 Cultural Dataset Construction

Training Data

- Knowledge Field Length Distribution



환경 및 데이터셋 State/Action/Reward

State

- q: query
- y_{<t}: generated tokens (y_0, … , y_{t-1})

$$s_t = \big(q,\ y_{<t}\big),$$
$$\mathcal{S} := \{\,(q, y_{<t})\,\}.$$

Action

- V:

$$\mathcal{A} := \mathcal{V},$$
$$a_t \in \mathcal{A} \quad (\text{next token } y_t),$$
$$\pi_\theta(a_t \mid s_t) = p_\theta\big(y_t = a_t \,\big|\, q, y_{<t}\big).$$

Reward

- Most

$$y_t^{(s)} \sim \pi_\theta(\cdot \mid s_t),$$
$$A_t = \begin{cases} 1, & \text{if teacher accepts } y_t^{(s)}, \\ 0, & \text{if teacher replaces it with } y_t^{(t)}. \end{cases}$$

Building a Cultural Context Aware AI

1. 개요
2. 환경 및 데이터셋
3. 알고리즘 소개
    1) LUPI
    2) SKD
4. 실험 설계
5. 결과 분석
6. 토의 및 결론

Vapnik, Vladimir, and Rauf Izmailov. "Learning using privileged information: Similarity control and knowledge transfer." *J. Mach. Learn. Res.* 16.1 (2015): 2023-2049.

알고리즘 소개 LUPI

Learning Using Privileged Information(LUPI)

- Training with privileged triplets: (x, x*, y) where x* is privileged information that only teacher knows
- Originally used in classification tasks

$$(x_1, x_1^*, y_1), \ldots, (x_\ell, x_\ell^*, y_\ell), \quad x_i \in X, \quad x_i^* \in X^*, \quad y_i \in \{-1, +1\},$$

LUPI triplet

- We adopt this triplet on our dataset, giving teacher query+knowledge but only query on student model
- M_t, M_s: Teacher / Student LM

$$(x, x^*, y) = (Q, K, Y), \quad Y = (R, A)$$

$$M_t(Y \mid Q, K)$$
$$M_s(Y \mid Q)$$

Xu, Wenda, et al. Speculative knowledge distillation: Bridging the teacher-student gap through interleaved sampling, ICLR'25

알고리즘 소개 SKD

Speculative Knowledge Distillation (SKD) (On-Policy RL used in KD field)

- For given prompt x and answer(reference) sequence y = (y_1, … ,y_{L_y}), compare two distribution for every decoding step i.

$$D(M_t||M_s)(y|x) = \frac{1}{L_y}\sum_{i=1}^{L_y} D(M_T(.|y_{<i},x)||M_s(.|y_{<i},x))$$

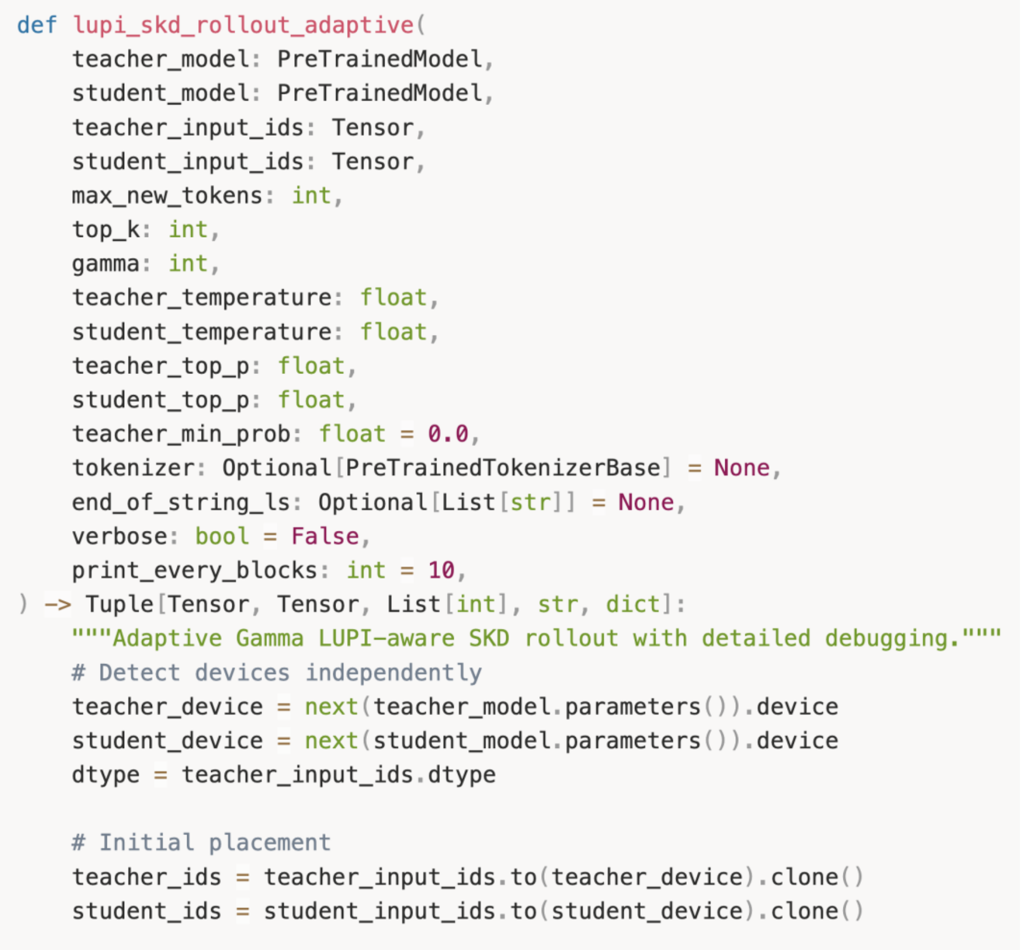
- Training objective – Make student model's distribution to be similar of teacher's one.

$$\min_{\theta_s} \; D(M_t\|M_s)(y \mid x)$$

알고리즘 소개 Codebase Examples

LUPI-aware SKD Rollout

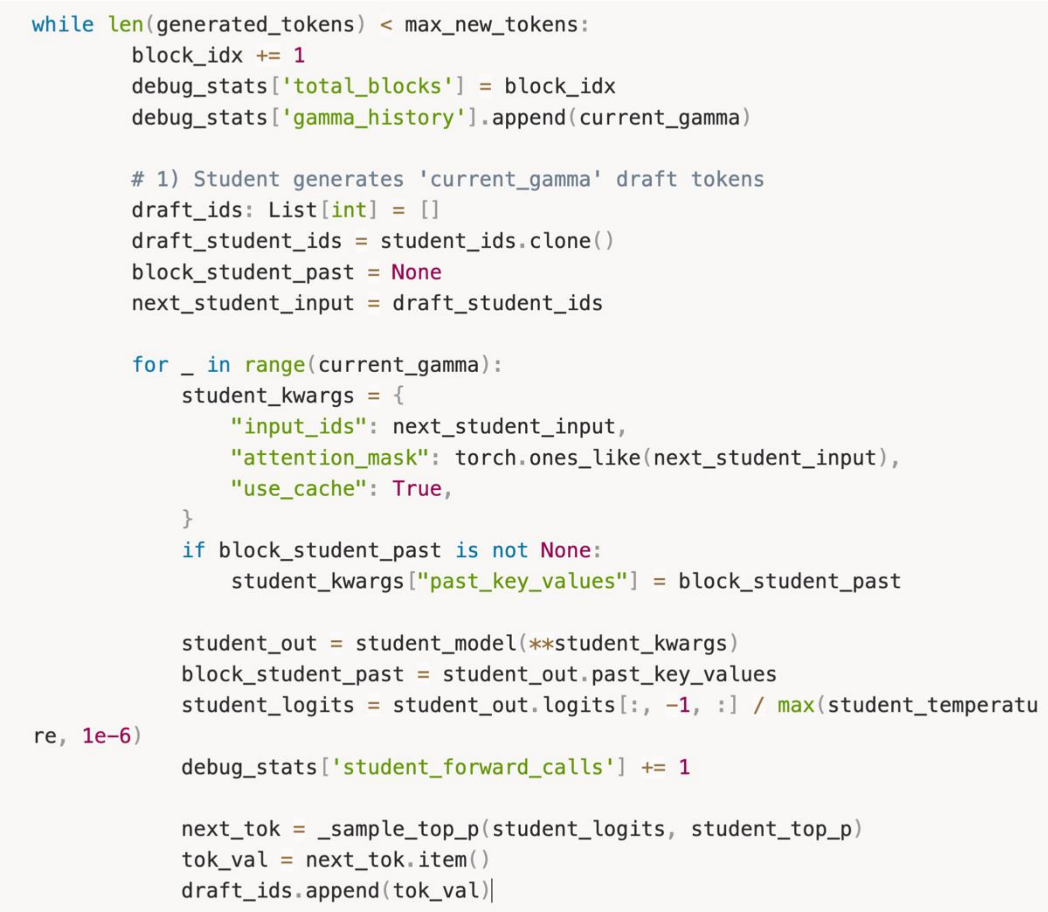
- teacher_input_ids are Query+ Knowledge
- student_input_ids are only Query



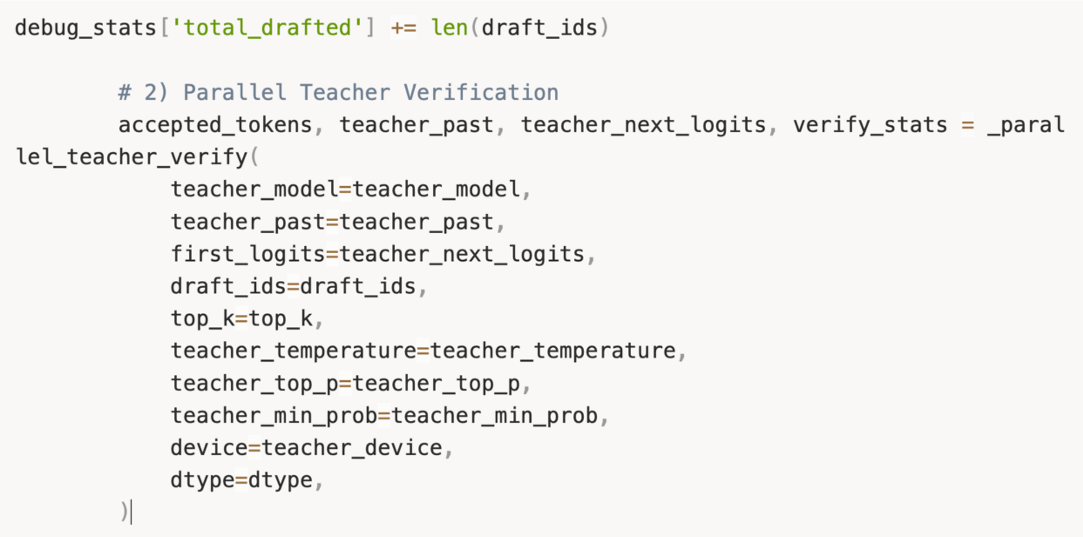
```python
def lupi_skd_rollout_adaptive(
    teacher_model: PreTrainedModel,
    student_model: PreTrainedModel,
    teacher_input_ids: Tensor,
    student_input_ids: Tensor,
    max_new_tokens: int,
    top_k: int,
    gamma: int,
    teacher_temperature: float,
    student_temperature: float,
    teacher_top_p: float,
    student_top_p: float,
    teacher_min_prob: float = 0.0,
    tokenizer: Optional[PreTrainedTokenizerBase] = None,
    end_of_string_ls: Optional[List[str]] = None,
    verbose: bool = False,
    print_every_blocks: int = 10,
) -> Tuple[Tensor, Tensor, List[int], str, dict]:
    """Adaptive Gamma LUPI-aware SKD rollout with detailed debugging."""
    # Detect devices independently
    teacher_device = next(teacher_model.parameters()).device
    student_device = next(student_model.parameters()).device
    dtype = teacher_input_ids.dtype

    # Initial placement
    teacher_ids = teacher_input_ids.to(teacher_device).clone()
    student_ids = student_input_ids.to(student_device).clone()
```

알고리즘 소개 Codebase Examples

Student draft with Parallel Teacher Verification (SKD loop)



```
while len(generated_tokens) < max_new_tokens:
        block_idx += 1
        debug_stats['total_blocks'] = block_idx
        debug_stats['gamma_history'].append(current_gamma)

        # 1) Student generates 'current_gamma' draft tokens
        draft_ids: List[int] = []
        draft_student_ids = student_ids.clone()
        block_student_past = None
        next_student_input = draft_student_ids

        for _ in range(current_gamma):
            student_kwargs = {
                "input_ids": next_student_input,
                "attention_mask": torch.ones_like(next_student_input),
                "use_cache": True,
            }
            if block_student_past is not None:
                student_kwargs["past_key_values"] = block_student_past

            student_out = student_model(**student_kwargs)
            block_student_past = student_out.past_key_values
            student_logits = student_out.logits[:, -1, :] / max(student_temperatu
re, 1e-6)
            debug_stats['student_forward_calls'] += 1

            next_tok = _sample_top_p(student_logits, student_top_p)
            tok_val = next_tok.item()
            draft_ids.append(tok_val)
```



```
debug_stats['total_drafted'] += len(draft_ids)

        # 2) Parallel Teacher Verification
        accepted_tokens, teacher_past, teacher_next_logits, verify_stats = _paral
lel_teacher_verify(
            teacher_model=teacher_model,
            teacher_past=teacher_past,
            first_logits=teacher_next_logits,
            draft_ids=draft_ids,
            top_k=top_k,
            teacher_temperature=teacher_temperature,
            teacher_top_p=teacher_top_p,
            teacher_min_prob=teacher_min_prob,
            device=teacher_device,
            dtype=dtype,
        )
```

draft_ids: action sequence that student proposed
top_k + min prob (top_p) verify with teacher

Building a Cultural Context Aware AI

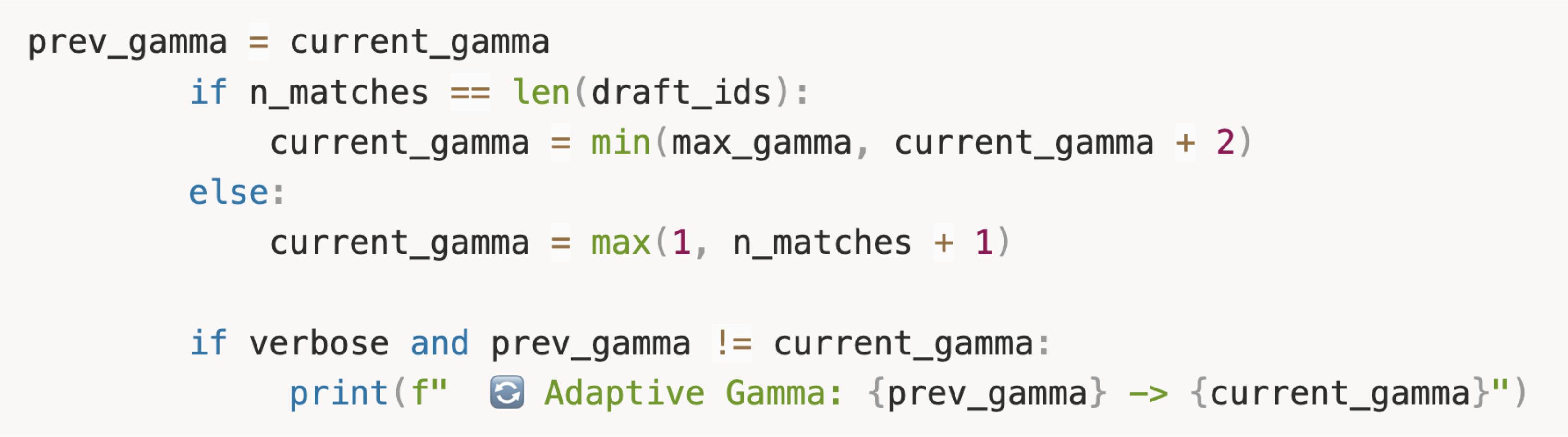
1. 개요
2. 환경 및 데이터셋
3. 알고리즘 소개
4. 실험 설계
    1) Baselines and Variants
    2) Experimental Setup
5. 결과 분석
6. 토의 및 결론

실험 설계 Baselines and Variants

Model Configuration

- Teacher: Qwen3-32B
- Student: Qwen3-4B (this model is originally distilled from Qwen3-32B)
  ➔ Same tokenizer / architecture family

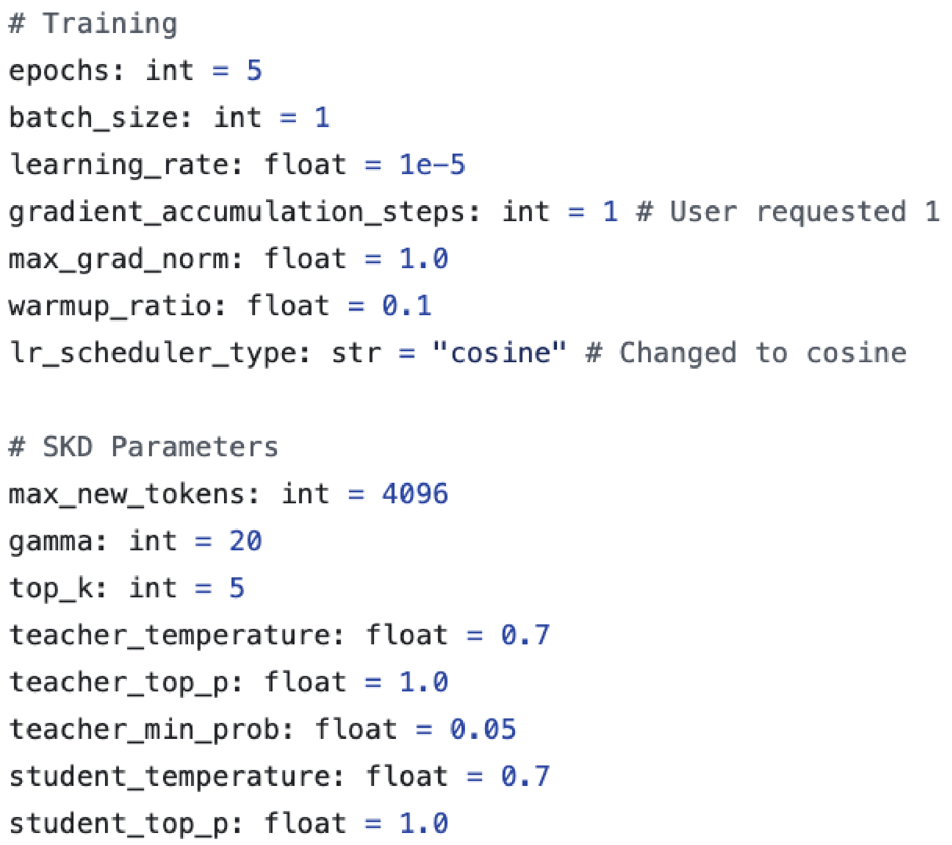
- Used adaptive gamma for dynamic block size (heuristic)
  ➔ n_matches: number of tokens where teacher model accepted
  ➔ If gamma increases, we think that the teacher trust the student more



```
prev_gamma = current_gamma
        if n_matches == len(draft_ids):
            current_gamma = min(max_gamma, current_gamma + 2)
        else:
            current_gamma = max(1, n_matches + 1)

        if verbose and prev_gamma != current_gamma:
             print(f"🔄 Adaptive Gamma: {prev_gamma} -> {current_gamma}")
```

실험 설계 Experimental Setup

Hyperparameters

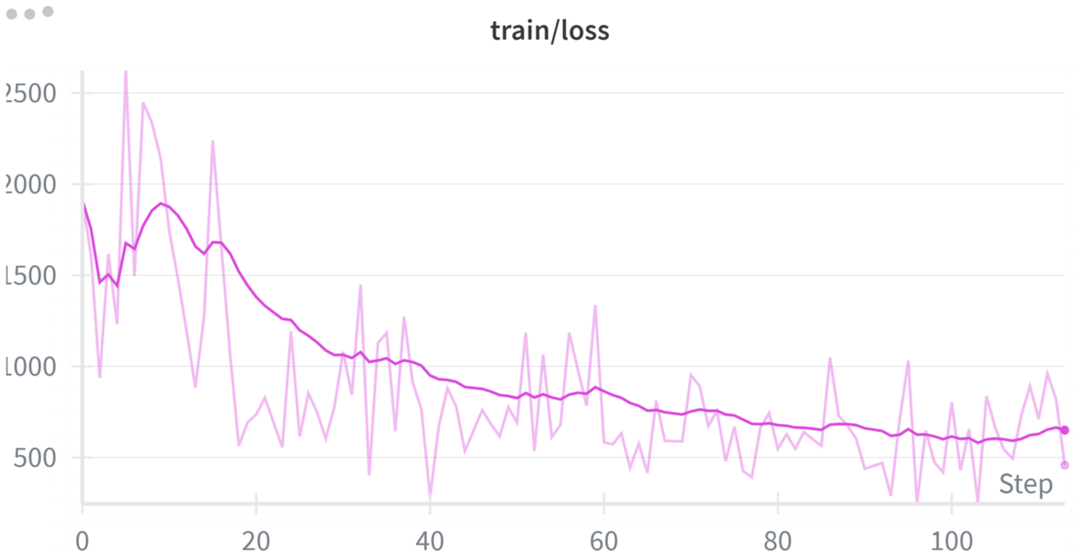


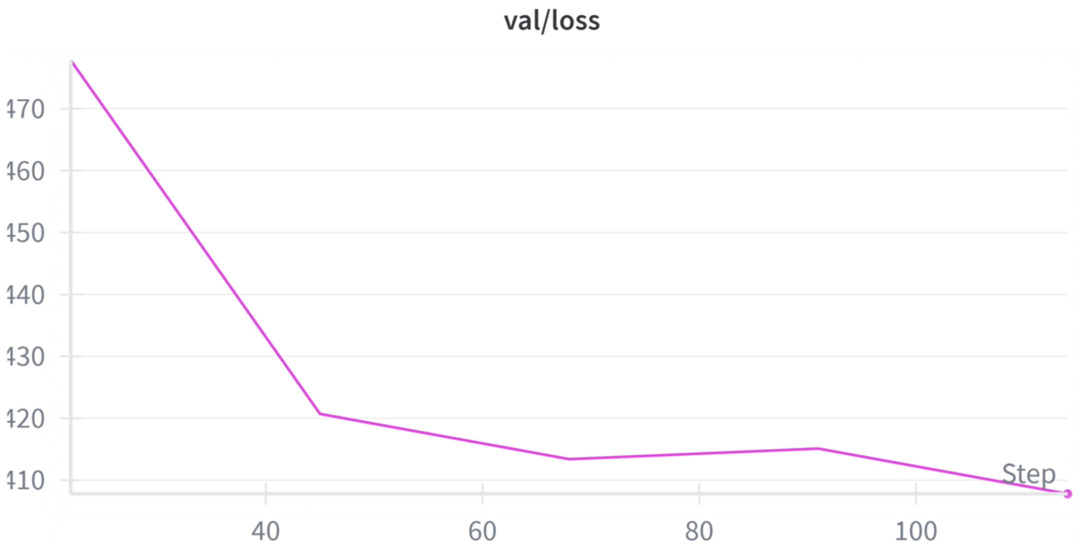
```
# Training
epochs: int = 5
batch_size: int = 1
learning_rate: float = 1e-5
gradient_accumulation_steps: int = 1 # User requested 1
max_grad_norm: float = 1.0
warmup_ratio: float = 0.1
lr_scheduler_type: str = "cosine" # Changed to cosine

# SKD Parameters
max_new_tokens: int = 4096
gamma: int = 20
top_k: int = 5
teacher_temperature: float = 0.7
teacher_top_p: float = 1.0
teacher_min_prob: float = 0.05
student_temperature: float = 0.7
student_top_p: float = 1.0
```

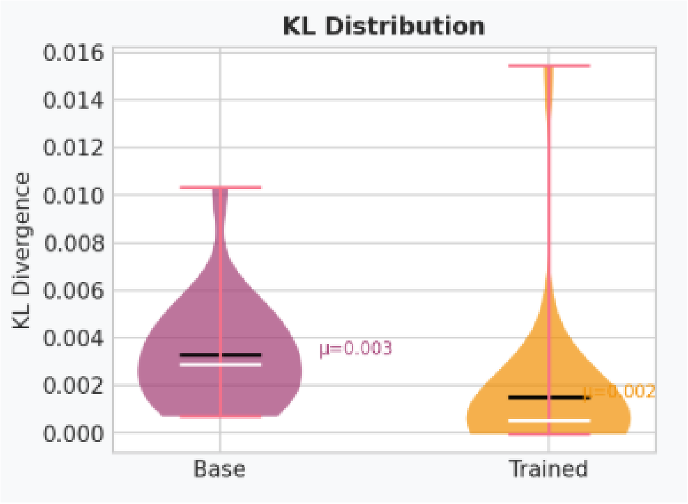
Building a Cultural Context Aware AI

1. 개요
2. 환경 및 데이터셋
3. 알고리즘 소개
4. 실험 설계
5. 결과 분석
   1) Quantitative Results
   2) Ablation Study
6. 토의 및 결론

결과 분석 Quantitative Results







결과 분석 Quantitative Results

➔ Input Query: How did Choi Chung-heon take power during the Goryeo military regime and what were the political reforms he implemented?

Qwen3-4B (Baseline)	Qwen3-4B (LUPI-SKD)
la información que has proporcionado sugiere que hay un error en la pregunta. Choi Chung-heon no fue un líder durante el régimen militar de Goryeo. El período de Goryeo (918-1392) fue una dinastía coreana que se dio en el sur de Corea, y no hay registros históricos de un líder llamado Choi Chung-heon en ese período. En realidad, en la historia coreana, hay un personaje llamado Choi Chung-heon, pero fue un líder durante la dinastía Joseon (1392-1897), no Goryeo. … Your question contains a historical inaccuracy. There is no historical figure named Choi Chung-heon who took power during the Goryeo military regime (a misnomer, as Goryeo was never a "military regime" in the modern sense).	the user is asking about Choi Chung-heon's rise to power and his political reforms during the Goryeo military regime. First, I need to confirm the background knowledge provided. The key points from the background are: Choi Chung-heon was a general who took power in 1056 by overthrowing King Taejong and King Taejo (Taejo). He established the military regime, which is considered a "military coup." His political reforms include: … Choi Chung-heon's rise to power and political reforms during the Goryeo military regime can be understood as follows: How He Took Power Choi Chung-heon, a military leader, seized control of the Goryeo royal court in 1056 by overthrowing King Taejong (the 26th king of Goryeo) and King Taejo (the 27th king).

➔ Qwen3-4B model makes hallucination(even Spanish in reasoning path!) while our trained model generates correct answer

Building a Cultural Context Aware AI

1. 개요
2. 환경 및 데이터셋
3. 알고리즘 소개
4. 실험 설계
5. 결과 분석
6. 토의 및 결론
    1) Limitations
    2) Conclusion/Future Work

토의 및 결론 Limitations

Language Limitations

- Due to the lack of Korean reasoning abilities on Qwen models, all experiments are conducted on the English(translated version) of the cultural dataset

Out of Domain Benchmark Evaluations

- We did not evaluate the trained student model on cultural benchmarks (e.g., CLIcK, BLEnD).

General Reasoning Ablilties

- We did not systematically measure whether injecting cultural knowledge via LUPI + SKD degrades general reasoning or standard NLP performance, leaving the trade-off between cultural specialization and overall capability unquantified.

토의 및 결론 Conclusion/Future Work

Conclusion

- With using small data (~200 samples, 180 are train and 20 are validation set), the model can generate reasoning paths about cultural reasoning.

토의 및 결론 Conclusion/Future Work

Future Works

Adaptive Top-K for given query

Mixing with General Reasoning data (Ablation Study)

- While injecting cultural knowledge in small LRMs, degregation of general reasoning ability can easily observed.
- Mixing the data with general reasoning (Math, code, commonsense) is necessary.
  - 7:3, 5:5, 3:7, two step training with cultural reasoning and general reasoning ablation will be performed.

➔ All future works are written in future_work.md

Appendix Contributors

박현빈: LUPI, SKD 구현 및 학습, 최적화

이정빈: Cultural Dataset 생성, 코드 배포 및 문서 작성