

학부생 연구실 인턴 활동보고서

인적사항	소속학과	수학과	성명	김연범
	학 번	202000826	학 년	4학년
전체 활동 기간	2025. 3. 1. ~ 2025. 8. 31 (6개월)			
해당월 활동 기간	2025. 5. 1. ~ 2025. 5. 31 (1개월)			
활동 내용	<ul style="list-style-type: none"> - LoRA 논문 탐독 및 정리 (LoRA: Low-Rank Adaptation Of Large Language Models) - 삼성 프로젝트: Decision Tree+RAG 기반 LLM 코드 인스펙션 <ol style="list-style-type: none"> 1) 수신된 메시지를 STT를 활용해 LLM input으로 변환 2) RAG + Decision Tree를 활용해 intent 파악 3) LLM output + TTS를 활용해 음성 메시지 생성 및 송신 			
활동 성과 및 소감	<p>a) 활동 성과</p> <ul style="list-style-type: none"> - LoRA 논문 탐독을 통한 LLM 구조에 대한 심층적 이해 달성 및 Low-Rank 개념 이해 달성. 나아가 Adapter 기반 LLM fine-tuning 방법에 대한 이해 달성 - RAG의 효용성을 위한 Decision Tree 기반 Data Embedding 과정에 대한 이해 달성 - konlp의 Kkma를 활용한 한국어 형태소 분할 방법에 대한 이해 및 open-weights 언어 모델인 Llama 사용 방법에 대한 이해 달성 <p>b) 소감</p> <ul style="list-style-type: none"> - LLM을 각 산업 분야에 활용하기 위해서 사용되는 다양한 기술들을 알게 돼 인상적이었다. - RAG 파이프라인에서 Decision Tree를 활용해 Retriever 성능을 향상시키는 idea가 인상적이었다. 			
평가	성실성	업무능력	활용효과	
	(상). 중 . 하	(상). 중 . 하	(상). 중 . 하	
* 활동사진 첨부 2025년 5월 21일 보고자 : 김연범 김연범 담당교수(지도교수) : 김경식 김경식				
충남대학교 소프트웨어중심대학사업단장 귀하				

활동사진(2장)

3. UNBALANCING THE LAW BY DISPROPORTION

Given the unusual advantages of *LobA*, we hope to further explore the properties of the low-methionine form from diverse sources. Note that the low-methionine form only lowers the threshold because it may which allows *in-vitro* multiple exposures in parallel. But also gives better irreversibility of form: the uptake samples are correlated with the permanent weight. We have run mostly on GRHL-1588, where we achieved the largest reduction of translatable proteins.

We provide a sequence of prompted inputs, to answer the following questions: 1) Given a parameter budget constraint, which subset of words matters in a p-locally Tied-Distinct? Should we add:

to maintain distinctive performance? 2) Is the "spared" adaptation metric ΔH^* really such a "spare"? That is, what is a good task to use in practice? 3) What is the connection between ΔH^* and H^* ? Does ΔH^* implicitly correlate with H^* ? How large is ΔH^* compared to H^* ?

We believe that our analysis is important for (3) and (4) shed light on the fundamental principles of using the learned common models for downstream tasks, such as classification and regression.

При этом Генрих Бендер не допускает возможности, в которой бы, в зависимости от того

7.3 WIECH WILNYI MATERIAŁY I TECHNIKI SĄSIADOWEJ LITERATURY

Given a fixed pattern of Budget, which types of weights should we assign with LRNN?

We focus on performance on distributed tasks.¹⁰ As mentioned in Section 4.2, we only consider weight matrices in the full attention module. We set a parameter budget of 1004 (roughly 15MB) if stored in FP16 or 2075.7 (31B), which corresponds to $\gamma = 5$; if we adopt one type of memory weights or $\gamma = 4.6$ for adapt strategies for all the layers. The result is presented in Table 5.

	1	2	3	4	5	6
2000-2001	70.1	70.6	71.2	71.3	71.4	71.5

Table 3: Validation accuracy on WikiQ2 and Multi30k when applying LReLU to different types of attribute weights in GPT-2, given the size number of trainable parameters. Among both H_1 and H_2 , LReLU with H_1 achieves better validation accuracy.

11. give the best performance overall. We find the standard deviation across random seeds to be consistent with our theoretical analysis.

Note that passing all the parameters in SST_1 or SST_2 leads to significantly lower performance, while setting both T_1 and T_2 to 0 yields the best result. The parameter α controls the trade-off between

Table 5: Validation accuracy on BioNLP and MatixNLP after applying L1/L2 to different types of attention weights in GPT-2 given the same number of trainable parameters. Adapting both H_1 and H_2 give the best performance overall. We used the standard deviation across random seeds to be conservative, but see the Additional file 1 for more details.

Note that passing all the parameters in SPL_{I} or SPL_{II} , results in significantly lower performance, while selecting both SPL_{I} and SPL_{II} results the best result. The parameter α controls the trade-off between the two models.

• 14

LeRaG가 경험적으로 뛰어난 부수를 보이는 편을, 이 논증은 디스크로프 커플에서 한층 더 low-rail orientation 특성을 더 잘 살펴보라고 한다.

총리 100-100% 시장에서는 다음과 같은 두 가지 유형의 이점을 제공하는데

LoRaWAN 기준 가중치 범위를 고려한 네트워크는 것이 아니며, 특정 고유자 생활에 대처 과정(leave-grade) 또는 생활을 추적하는 범위이다. 하지만 모든 가중치 범위를 충분하게 조작할 수 있는 것은 아니며, 일부 범위는 고정되어 있다.

(2) 대안적 품종 개발 또는 실제로 저수지임기?)
LORA는 저수지위 멜็ด를 이 물위에 가중치를 조절하는 기름이지기 때문에 “신경적” 즉 저 수위 멜็ด 저수지 위에 설치 *break-off device*를 가지는 경우는 허용하지 않았다. 한편 모두 같은 저수지 위

9) 본문에서는 실업을 통해 실제 과정에서 필요한 최소한의 행동을 찾고, LoRa의 성능과 효율성을 극대화할 수 있는 헨크 예술을 분석한다.

(2) The term "operator" means a person or entity that provides

