ChatGPT 相关技术培训

培训单位：东南大学-KSE 实验室

指导教师：漆桂林教授

方案撰写人：毕胜、陈永锐、翟松林

Contents

1. Introduction

[Recommended Reading]

- Human Language Understanding & Reasoning
- Attention Is All You Need (Transformers)
- Blog Post: The Illustrated Transformer
- HuggingFace's course on Transformers

What are LLMs?

2. BERT (encoder-only models)

(BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding)

[Recommended Reading]

- Deep contextualized word representations (ELMo)
- Improving Language Understanding by Generative Pre-Training (OpenAI GPT)
- RoBERTa: A Robustly Optimized BERT Pretraining Approach
- ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators

3. T5 (encoder-decoder models)

(Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer (T5))

[Recommended Reading]

- Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer (T5)
- BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension
- mT5: A massively multilingual pre-trained text-to-text transformer
- AlexaTM 20B: Few-Shot Learning Using a Large-Scale Multilingual Seq2Seq Model

4. GPT-3 (decoder-only models)

(Language Models are Few-Shot Learners (GPT-3))

[Recommended Reading]

- Language Models are Unsupervised Multitask Learners (GPT-2)
- PaLM: Scaling Language Modeling with Pathways
- OPT: Open Pre-trained Transformer Language Models

How to Use and Adapt LLMs?

5. Prompting for few-shot learning

Making Pre-trained Language Models Better Few-shot Learners (blog post)

How Many Data Points is a Prompt Worth?

[Recommended Reading]

- Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference
- True Few-Shot Learning with Language Models
- Cutting Down on Prompts and Parameters: Simple Few-Shot Learning with Language Models
- Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing

6. Prompting as parameter-efficient fine-tuning

Prefix-Tuning: Optimizing Continuous Prompts for Generation

The Power of Scale for Parameter-Efficient Prompt Tuning

[Recommended Reading]

- Factual Probing Is [MASK]: Learning vs. Learning to Recall
- P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Universally Across Scales and Tasks
- LoRA: Low-Rank Adaptation of Large Language Models
- Towards a Unified View of Parameter-Efficient Transfer Learning

7. In-context learning

Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?
An Explanation of In-context Learning as Implicit Bayesian Inference (we don't expect you to read this paper in depth, you can check out this blog post instead)

[Recommended Reading]

- What Makes Good In-Context Examples for GPT-3?
- Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity
- Data Distributional Properties Drive Emergent In-Context Learning in Transformers
- What Can Transformers Learn In-Context? A Case Study of Simple Function Classes

8. Calibration of prompting LLMs

Calibrate Before Use: Improving Few-Shot Performance of Language Models
Surface Form Competition: Why the Highest Probability Answer Isn't Always Right

[Recommended Reading]

- Noisy Channel Language Model Prompting for Few-Shot Text Classification
- How Can We Know When Language Models Know? On the Calibration of Language Models for Question Answering
- Language Models (Mostly) Know What They Know

9. Reasoning

Chain of Thought Prompting Elicits Reasoning in Large Language Models
Large Language Models are Zero-Shot Reasoners

[Recommended Reading]

- Explaining Answers with Entailment Trees
- Self-Consistency Improves Chain of Thought Reasoning in Language Models
- Faithful Reasoning Using Large Language Models

10. Knowledge

Language Models as Knowledge Bases?

How Much Knowledge Can You Pack Into the Parameters of a Language Model?

[Recommended Reading]

- Knowledge Neurons in Pretrained Transformers
- Fast Model Editing at Scale
- Question and Answer Test-Train Overlap in Open-Domain Question Answering Datasets

Dissecting LLMs: Data, Model Scaling and Risks

11. Data

Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus

[Recommended Reading]

- The Pile: An 800GB Dataset of Diverse Text for Language Modeling
- Deduplicating Training Data Makes Language Models Better

12. Scaling

Training Compute-Optimal Large Language Models

[Recommended Reading]

- Scaling Laws for Neural Language Models
- Scale Efficiently: Insights from Pre-training and Fine-tuning Transformers
- Scaling Laws for Autoregressive Generative Modeling

13. Privacy

Extracting Training Data from Large Language Models

[Recommended Reading]

- Quantifying Memorization Across Neural Language Models
- Deduplicating Training Data Mitigates Privacy Risks in Language Models
- Large Language Models Can Be Strong Differentially Private Learners
- Recovering Private Text in Federated Learning of Language Models

14. Bias & Toxicity I- evaluation

RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models

[Recommended Reading]

- On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
- Red Teaming Language Models with Language Models
- Whose Language Counts as High Quality? Measuring Language Ideologies in Text Data Selection

15. Bias & Toxicity II- mitigation

Self-Diagnosis and Self-Debiasing: A Proposal for Reducing Corpus-Based Bias in NLP

[Recommended Reading]

- Challenges in Detoxifying Language Models
- Detoxifying Language Models Risks Marginalizing Minority Voices
- Plug and Play Language Models: A Simple Approach to Controlled Text Generation
- GeDi: Generative discriminator guided sequence generation

Beyond Current LLMs: Models and Applications

16. Sparse models

Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity

[Recommended Reading]

- Efficient Large Scale Language Modeling with Mixtures of Experts
- Branch-Train-Merge: Embarrassingly Parallel Training of Expert Language Models
- A Review of Sparse Expert Models in Deep Learning

17. Retrieval-based LMs

Improving language models by retrieving from trillions of tokens

- Generalization through Memorization: Nearest Neighbor Language Models
- Training Language Models with Memory Augmentation
- Few-shot Learning with Retrieval Augmented Language Models

18. Training LMs with human feedback

Training language models to follow instructions with human feedback

[Recommended Reading]

- Learning to summarize from human feedback
- Fine-Tuning Language Models from Human Preferences
- MemPrompt: Memory-assisted Prompt Editing with User Feedback
- LaMDA: Language Models for Dialog Application

19. Code LMs

Evaluating Large Language Models Trained on Code

[Recommended Reading]

- A Conversational Paradigm for Program Synthesis
- InCoder: A Generative Model for Code Infilling and Synthesis
- A Systematic Evaluation of Large Language Models of Code
- Language Models of Code are Few-Shot Commonsense Learners
- Competition-Level Code Generation with AlphaCode

20. Multimodal LMs

Flamingo: a Visual Language Model for Few-Shot Learning

[Recommended Reading]

- Blog post: Generalized Visual Language Models
- Learning Transferable Visual Models From Natural Language Supervision (CLIP)
- Multimodal Few-Shot Learning with Frozen Language Models
- CM3: A Causal Masked Multimodal Model of the Internet

21. AI Alignment + open discussion

[Recommended Reading]

- A General Language Assistant as a Laboratory for Alignment
- Alignment of Language Agents
- Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback

References

https://www.cs.princeton.edu/courses/archive/fall22/cos597G/

https://github.com/KSESEU/LLMPapers