236004: Transformers and Attention-based Networks

Winter 2023-24

Staff

Instructor	Office location & hours
Gilad Levy	TBA
Assaf Schuster	TBA
Yonatan Belinkov	TBA

General information

Description

Since their first appearance back in 2017 in the paper "Attention is all you need" Vaswani et al. 2017, Transformers became the standard model for almost every task in Deep Learning.
This course will cover the components that the transformers is comprised of and the part each one plays in the complete function of the model as well as advanced/SOTA Transformer-based methods for various tasks.

Expectations and goals

1. Review the fundamentals of Deep Learning and explain the different pros and cons of different models in the context of the Transformer and Attention in general.
2. Train various Deep neural networks and analyze the various challenges that arise when using them.
3. Construct/Modify Transformer-based networks and apply them on different domains.
4. Explain the reasons for success/failure in a network from a more theoretical point of view while utilizing fundamental concepts such as manifolds.

Grading

Students will be graded based mainly on a final project as well as several short assignments.

Course materials

Required materials

- 236756 – Intro to ML

And at least one of the following courses:

- 236781 - Deep learning on computational accelerators
- 236299 – Natural language processing

Optional materials

This is an advanced course in Deep Learning which assumes an understanding of the fundamentals of deep learning and as such the course 236781- Deep learning on computational accelerators is recommended. Additionally, familiarity with Python and Pytorch specifically is assumed as we will rely heavily on Pytorch based models and implementations.

Course schedule (Order TBA)

Week	Topic	Reading	Exercises
Week 1	Attention	Attention is all you need	TBA
Week 2	Transformers	Attention is all you need	TBA
Week 3	Encoder-Decoder models	Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer	TBA
Week 4	Generative models	Language Models are Few-Shot Learners	TBA
Week 5	In-context learning	How does in-context learning work? A framework for understanding the differences from traditional supervised learning	TBA
Week 6	Encoder-only models	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	TBA
Week 7	Vision	An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale	TBA
Week 8	Multi-modality	Learning Transferable Visual Models From Natural Language Supervision	TBA
Week 9	Sparsity, architecture and scaling	1. FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness 2. Switch Transformers: Scaling to Trillion	TBA

Week	Topic	Reading	Exercises
		Parameter Models with Simple and Efficient Sparsity	
Week 10	Expressivity and interpretability	TBA	TBA
Week 11	RL & transformers	1. Reinforcement Learning from Human Feedback 2. Decision Transformer: Reinforcement Learning via Sequence Modeling	TBA
Week 12	Projects		
Week 13	Projects		

Additional information and resources

Reading materials

- "Attention is all you need" Vaswani et al. 2017 is the fundamental paper which introduced transformers and we highly recommend reading it thoroughly.
- The Illustrated Transformer – A carefully constructed blog which goes into the full details of the Transformer with many helpful visuals.