

DOKTORANTŪROS STUDIJŲ DALYKO SANDAS

Dalyko pavadinimas	Mokslo kryptis, kodas	Fakultetas	Katedra
Gilieji neuroniniai tinklai	Informatika (N 009), Informatikos inžinerija (T 007)	MIF	Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų būdas	Kreditų skaičius ECTS	Studijų būdas	Kreditų skaičius
paskaitos	1 (pavasario)	konsultacijos	1
individualus	4	seminarai	1

Dalyko anotacija

Nors dirbtiniai neuroniniai tinklai pradėti kurti prieš daugiau nei 50 metų, pastarąjį dešimtmetį įvyko esminis lūžis dėl naujų architektūrų (Transformerių, Difuzijos modelių) ir išaugusių skaičiavimo resursų. Šio kurso metu doktorantai susipažins ne tik su klasikineis metodais, bet ir su šiuolaikinėmis giliojo mokymosi (Deep Learning) technologijomis. Didžiausias dėmesys skiriamas architektūroms, leidžiančioms spręsti sudėtingus generatyvinius, natūralios kalbos apdorojimo (NLP) ir vaizdo analizės uždavinius.

☐ Giliojo mokymosi pagrindai ir optimizavimas:

- Neuroninių tinklų teorija: Biologinis įkvėpimas ir matematinis formulavimas. Universalios aproksimacijos teorema.
- Pažangus optimizavimas: Stochastinis gradientinis nusileidimas (SGD) ir adaptyvūs metodai (Adam, AdamW, RMSprop). Tikslų funkcijų paviršiai (Loss landscapes), balno taškai ir lokalūs minimumai.
- Regularizacija ir normalizavimas: „Dropout“, žymių glodinimas (Label Smoothing), svorių silpninimas (Weight Decay). Vidinio kovariacinio poslinkio problema: Paketo (Batch), Sluoksnių (Layer) ir Grupės (Group) normalizavimas, RMSNorm.
- Inicializavimo strategijos: Xavier (Glorot) ir He inicializavimas.

☐ Kompiuterinė rega ir konvoliucinės architektūros:

- CNN veikimo principai: Konvoliucijos (1D, 2D, 3D), išplėstinės (dilated) konvoliucijos, sutelkimo (pooling) sluoksniai ir žingsniai (strides). Receptyviniai laukai.
- Architektūrų evoliucija: Nuo LeNet/AlexNet iki ResNet (liekamieji ryšiai), DenseNet ir EfficientNet (sudėtinis mastelio keitimas).
- Moderni kompiuterinė rega: ConvNeXt (CNN modernizavimas pagal Transformerių principus).
- Objektų aptikimas ir segmentavimas: R-CNN šeima, YOLO (You Only Look Once), U-Net architektūros medicininiais vaizdams.

☐ Sekų modeliavimas ir Dėmesio mechanizmai:

- Rekurentiniai tinklai (Trumpa apžvalga): RNN, LSTM, GRU. Nykstančio gradiento problema ir atgalinis sklaidimas laike (BPTT).
- Dėmesio (Attention) mechanizmų revoliucija: Skaliamosios sandaugos dėmesio (Scaled Dot-Product Attention) matematika. Raktų, Užklausų ir Reikšmių (Key, Query, Value) matricos. Daugiagalvis dėmesys (Multi-Head Attention).
- Pozicinis kodavimas: Absoliutus (Sinusoidinis), apmokomas ir santykinis (RoPE - Rotary Positional Embeddings).

□ Transformeriai ir Didieji kalbos modeliai (LLM):

- Transformerių architektūros:
 - Tik kodavimo (Encoder-only): BERT, RoBERTa (Maskuotas kalbos modeliavimas klasifikavimui/vektORIZAVIMUI).
 - Tik dekodavimo (Decoder-only): GPT šeima, LLaMA (Priežastinis kalbos modeliavimas generavimui).
 - Kodavimo-Dekodavimo (Encoder-Decoder): T5, BART (Seq2Seq uždaviniai).
- Mokymo paradigmos: Išankstinis mokymas (Pre-training) ir pritaikymas (Fine-tuning).
- Suderinimas ir saugumas: Sustiprintas mokymasis iš žmogaus grįžtamojo ryšio (RLHF), tiesioginis preferencijų optimizavimas (DPO).
- Efektyvumas: Tokenizacija (BPE, WordPiece). Konteksto lango valdymas (Flash Attention).

□ Efektyvus parametrų pritaikymas (PEFT):

- Didelių modelių adaptavimas turint ribotus resursus.
- Technikos: LoRA (Low-Rank Adaptation), QLoRA, Prefix Tuning, P-Tuning.
- Kvantavimas: FP16, BF16, INT8 ir FP4 tikslumo mokymas.

□ Gilieji generatyviniai modeliai:

- Autoenkoderiai (AE): Informacijos suspaudimas („bottleneck“), triukšmą šalinantys AE.
- Variaciniai autoenkoderiai (VAE): Tikimybiniai pagrindai, ELBO optimizavimas, reparameterizacijos triukas. β -VAE ir latentinių požymių atskyrimas.
- Generatyviniai priešininkų tinklai (GAN): Minimax lošimas, Nasho pusiausvyra. „Mode collapse“ problema. WGAN ir gradiento bauda.
- Difuzijos modeliai (Naujasis standartas):
 - DDPM (Denoising Diffusion Probabilistic Models): Tiesioginė difuzija (Gauso triukšmo pridėjimas) ir atvirkštinė difuzija (triukšmo šalinimas per U-Net).
 - Latentiniai difuzijos modeliai (LDM): Generavimas latentinėje erdvėje (pvz., Stable Diffusion).
 - *Imties sudarymas (Sampling)*: DDIM, Euler Ancestral samplers. Generavimo valdymas: ControlNet, tekstiniai nurodymai (conditioning).

□ Multimodalinis mokymasis ir Grafų tinklai:

- Vaizdo-Kalbos modeliai: CLIP (Contrastive Language-Image Pre-training), SigLIP. „Zero-shot“ klasifikavimas.

- Grafų neuroniniai tinklai (GNN): Grafų konvoliuciniai tinklai (GCN), Grafų dėmesio tinklai (GAT). Pranešimų perdavimo (Message passing) sistema.

Išspręsti tyrimo uždavinį pasitelkus šiuolaikinius giliojo mokymo metodus (pvz., sukurti Transformerių modelį, pritaikyti LLM specifinei užduočiai arba apmokyti generatyvinį modelį).

Pagrindinė literatūra

1. Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola, "Dive into Deep Learning", 2023. <https://d2l.ai/>
2. Ian Goodfellow, Yoshua Bengio and Aaron Courville, "Deep Learning", MIT Press, 2016. <http://www.deeplearningbook.org>
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems. <https://arxiv.org/abs/1706.03762>
4. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In Advances in neural information processing systems. <https://arxiv.org/abs/2006.11239>
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. ICLR. <https://arxiv.org/abs/2010.11929>
6. Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-rank adaptation of large language models. ICLR. <https://arxiv.org/abs/2106.09685>
7. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. <https://arxiv.org/abs/2112.10752>
8. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In International conference on machine learning. <https://arxiv.org/abs/2103.00020>
9. Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S. & Lerchner, A. (2016). beta-VAE: Learning basic visual concepts with a constrained variational framework. <https://openreview.net/forum?id=Sy2fzU9gl>
10. Karpathy, A. (2016). Deep reinforcement learning: Pong from pixels. <http://karpathy.github.io/2016/05/31/rl/>

Konsultuojančiųjų dėstytojų vardas, pavardė	Mokslo laipsnis	Svarbiausieji darbai mokslo kryptyje, paskelbti per pastaruosius 5 metus
Povilas Treigys	dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Povilas+Treigys
Jolita Bernatavičienė	dr.	https://elaba.mb.vu.lt/dmsti/?aut=Jolita+Bernatavi%C4%8Dien%C4%97
Viktor Medvedev	dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Viktor+Medvedev