

Nome primário: Tópicos em Engenharia de Sistemas e Computação

Nome Secundário: Modelos de Linguagem Neurais

Ementa: Modelos de Linguagem probabilísticos. Modelos de Linguagem Neurais baseados em redes recorrentes. Modelos de Linguagem Neurais baseados em arquiteturas Transformers. Arquiteturas de modelos neurais. Pre-treinamento e ajuste fino. Treinamento eficiente. Modelos de contexto longo. *Large Language Models*. Aplicações.

Conteúdo Programático:

Conceitos básicos: semântica distribucional e embeddings estáticos; Redes Neurais.

Modelos probabilísticos de Linguagem baseados em n-grams.

Modelos de Linguagem baseados em redes recorrentes: ElMo.

Arquitetura codificadora-decodificadora e mecanismos de atenção. Modelos baseados na arquitetura Transformer: BERT, GPT, T5, etc.

Treinamento de modelos neurais: pré-treinamento e ajuste fino, extração de atributos e treinamento continuado.

Modelos de contexto longo: Longformer e Unlimiformer.

Large Language Models.

Treinamento eficiente de modelos neurais: *zero-shot*, *one-shot*, *few-shot*; adaptadores, *prompt* e *prefix tuning*, *instruct tuning*.

Agentes de conversação.

Aplicações.

Bibliografia: Natural Language Processing with Transformers, Revised Edition, by Lewis Tunstall, Leandro von Werra, Thomas Wolf. Released May 2022. Publisher(s): O'Reilly Media, Inc. ISBN: 9781098136796

Artigos da ACL, EMLNLP, EACL, LREC, ICLR, Neurips, Computational Linguistics, Natural Language Engineering, Transactions on Natural Language Processing, entre outros.

Pré-requisito (desejado): Aprendizado de Máquina ou Inteligência Artificial ou Mineração de Dados.

Horário: 3as e 5as de 7h às 9h