

MQI2950 **Tópicos Especiais: Processamento de Linguagem Natural**

Carga Horaria Total: 45 horas (3-0-0)

Créditos: 03

EMENTA

O processamento de linguagem natural (PLN) ou linguística computacional é uma das tecnologias mais importantes da era da informação: pesquisa na web, publicidade, e-mails, atendimento ao cliente, tradução de idiomas, agentes virtuais, relatórios médicos, processamento de normas, processos jurídicos, etc. Neste curso, os alunos terão uma introdução ao Aprendizado Profundo para PLN e realizarão aplicações práticas por meio de aplicações de modelos pré-treinados e bibliotecas de PLN, e atividades Hands-on, utilizando ferramenta comercial (Tornado). O Tornado é uma ferramenta desenvolvida pelo ICA da PUC-Rio para o processamento automático de grandes volumes de documentos que transforma dados não-estruturados (word, pdf, email, pptx, excel, tabelas, equações e imagens) em dados estruturados ou semiestruturados para formar data lakes e posterior análise e busca, classificação, agrupamento e questioning & answering por modelos de linguagem natural. NLP tem aplicações em metrologia. Um exemplo são sistemas de informação focados na melhoria dos processos regulatórios, construídos a partir de ferramentas de aprendizado de máquina, visualização e de análise das informações provenientes de múltiplas entidades reguladoras, capazes de gerir e acompanhar o conjunto de legislações e normativas (leis, resoluções, normas, padrões e ofícios) com uma visão multinível, através de ferramentas de busca inteligente. O resultado é rapidez na adequação a novos requisitos regulatórios, redução de não conformidades e multas. O objetivo do curso é apresentar introduzir Redes Neurais de Aprendizado Profundo para PLN, incluindo: a preparação dados com Text Mining; análise de frequência de texto e representações por frequência Tf-idf e bag of words; criação e aplicações com Embeddings, modelos de representação, word2vector, doc2vector sequence2vector; desenvolvimento de aplicações em classificação, previsão e agrupamento de texto; construção de aplicações com modelos de Deep Learning e modelos específicos LSTM; geradores de texto chat box. Os temas redes neurais, algoritmos de aprendizado e redes de aprendizado profundo serão apresentados conceitual e operacionalmente com o objetivo de apenas revelar como funcionam as aplicações de PLN e viabilizar o emprego de modelos pré-treinados. Alunos com base científica em redes neurais de aprendizado profundo se distinguirão através de seus projetos finais.

Introdução à Inteligência Artificial, Natural Processing Language, Text Mining, Modelos de Representação de Contexto Embeddings, Análise não supervisionada de dados, Machine Learning Supervisionada, Introdução a Deep Learning, Modelos BERT, Semântica Computacional, Named Entity Recognition.

Conteúdo Programático

Programa da disciplina Introdução à Inteligência Artificial • Conceitos e definições • História • Machine Learning • Deep Learning • Aplicações Redes Neurais Artificiais • Definição e conceitos • Características básicas • Arquitetura • Perceptron • Multilayer Perceptron • Funções de Ativação • Tarefas de Aprendizado • Jupyter Notebook (Colab) • Prática (keras + tensorflow) Algoritmos de Aprendizado • Gradiente Descendente • Gradiente Descendente Estocástico • Gradiente Descendente Mini Batch • Convergência vs Generalização • Taxa de Aprendizado e Momento • Prática (keras + tensorflow) Introdução à NLP • Estado da arte em NLP • Aplicações na indústria • Características • Ferramentas para análise de Dados • Limpeza e preparação dados • Tokenização • Stopwords • Lemetização e Stemming • Análise de frequência WordCloud • Análise de Sentimentos (baseado em léxico) • Exercícios (NLTK) Text Mining • Modelos de Representação • Bag of words • TF-IDF • Similaridade do cosseno • Aplicações de similaridade • Recomendação de produtos • Similaridade de textos jurídicos • Bibliotecas e infraestrutura Nuvem • Exercícios tutorados

Modelos de Representação de Contexto Embeddings • Word2vector • Glove • Doc2Vector
• Aplicações com Embeddings e supervisionados • Modelos de contexto • Aplicações na indústria • Modelos dimensionais Análise não supervisionada de dados • Análise de tópicos LDA • T-SNE clusterização por vizinhança • Geração de gráficos de interpretação • PCA • Aplicações Hands-On Machine Learning Supervisionado • Classificação • Classificação de texto • Métricas • Avaliação Deep Learning • Redes Recorrentes • LSTM • Geração de texto Deep Learning • Modelos pré-treinados • Persistência de Dados • Sumarização de Texto Deep Learning • Transformers • BERT • GPT2 • Análise de sentimentos com Transformers • Introdução a Questioning & Answering Chatbots Named Entity Recognition (NER)

Avaliação

Provas Teórica e Prática, projeto final, produção e submissão de artigo

BIBLIOGRAFIA PRINCIPAL

Bibliografia segundo as normas da ABNT 1. Ian Goodfellow and Yoshua Bengio and Aaron (2016), Deep Learning MIT press book, <https://www.deeplearningbook.org/> 2. Jurafsky and Martin (2008), Speech and Language Processing, Prentice Hall, 2nd Edition - <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf> 3. Manning, Schütze and Raghavan (2009), An Introduction to Information Retrieval, Online edition (c), Cambridge UP, <https://docero.com.br/doc/es80vv5> 4. Morgan & Claypool, Neural Network methods for Natural Language Processing. Yoav Goldberg. 2017, 309 pages, ISBN-10 : 1627052984, ISBN-13 : 978- 1627052986 5. Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Hars-hit Surana, Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems. (Published on June 17, 2020)