SYNTHESE

Réalisé par : Aymane MAHRI

Encadré par : Pr.Lotfi EL AACHAK

During Lab 4, we delved into the realm of Natural Language Processing (NLP) with the goal of becoming proficient in using PyTorch for language modeling. Our journey began with the task of collecting text data from various Arabic websites on a specific topic. Each text was assigned a relevance score, providing a basis for dataset preparation.

Following data collection, we established a robust preprocessing pipeline involving tokenization, stemming, lemmatization, and stop words removal. This ensured that our textual data was cleaned and prepared for analysis. We then ventured into training models using RNN, Bidirectional RNN, GRU, and LSTM architectures, fine-tuning hyperparameters along the way to optimize performance.

Transitioning to the realm of text generation, we installed pytorch-transformers and fine-tuned the GPT2 pre-trained model on a customized dataset. This allowed us to generate new paragraphs based on given sentences, expanding our understanding of text generation techniques.

Next, we turned our attention to BERT, utilizing the pre-trained bert-base-uncased model. We prepared the data and adapted the Bert Embedding Layer before fine-tuning and training the model with carefully selected hyperparameters. Evaluation metrics such as Accuracy, Loss, and F1 score, provided insights into the efficiency of our model.

Throughout the lab, we gained a comprehensive understanding of NLP techniques and their applications in machine learning tasks. We explored various word embedding strategies, honing our skills in preprocessing, encoding, and modeling textual data effectively. Leveraging tools such as spaCy, NLTK,Tensorflow; and PyTorch, we successfully applied different models and techniques to real-world datasets, enhancing our proficiency in NLP and gaining practical experience in handling regression and classification problems.

In conclusion, Lab 4 equipped us with valuable insights into the intricacies of NLP and machine learning, empowering us to tackle complex textual data analysis tasks with confidence. Through hands-on experimentation and exploration, we expanded our knowledge and skills, paving the way for future advancements in the field.